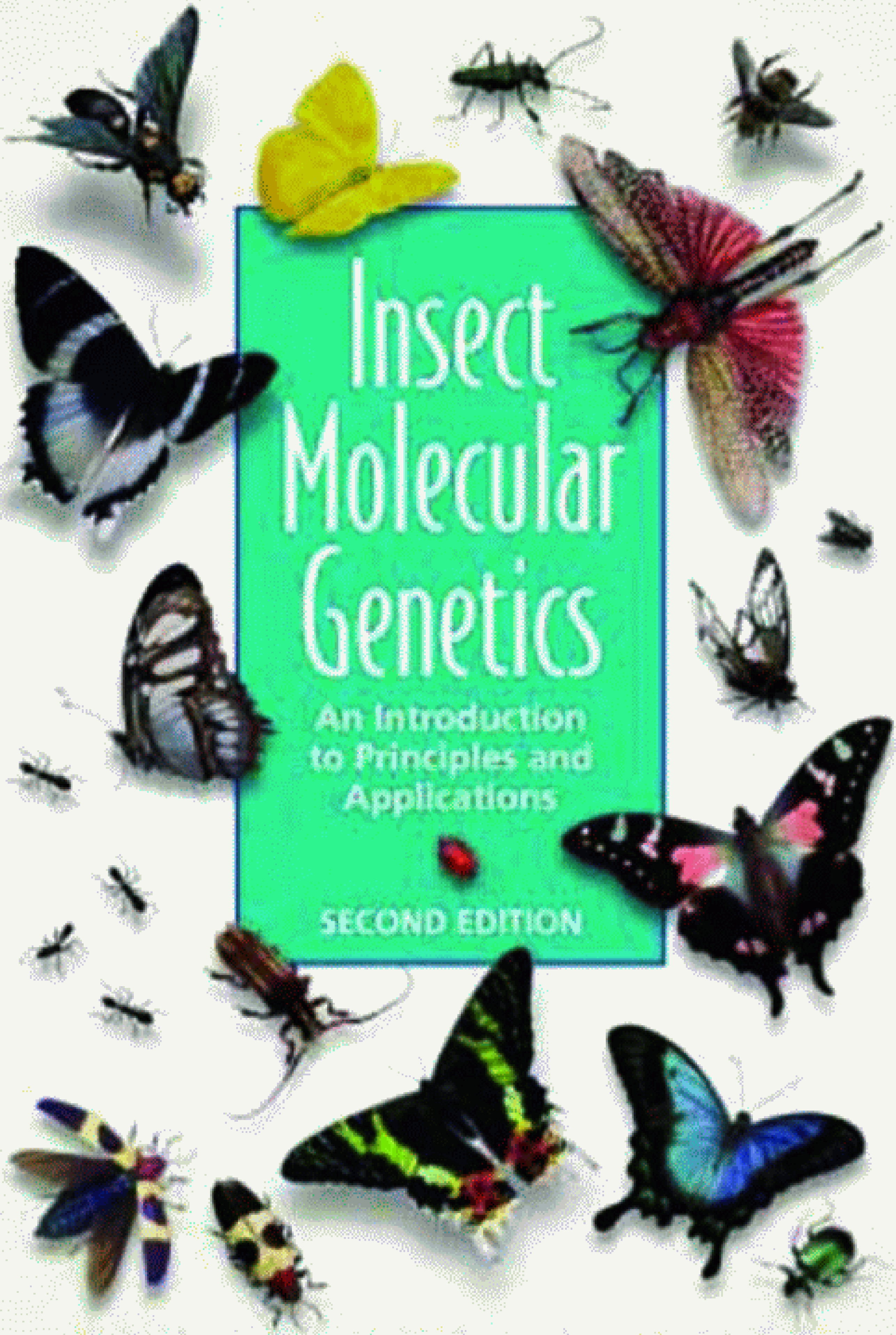


Insect Molecular Genetics

An Introduction to Principles and Applications

SECOND EDITION

MARJORIE A. HOY

This book is printed on acid-free paper. ♾

Academic Press
An imprint of Elsevier Science
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.academicpress.com

Academic Press
An imprint of Elsevier Science
200 Wheeler Road, Burlington, Massachusetts 01803, USA
http://www.academicpressbooks.com

Academic Press
84 Theobald's Road, London WC1X 8RR, UK
http://www.academicpress.com

Library of Congress Control Number: 2002110815

International Standard Book Number: 0-12-357031-X

PRINTED IN THE UNITED STATES OF AMERICA
02 03 04 05 06 9 8 7 6 5 4 3 2 1

CONTENTS

PART II: MOLECULAR GENETIC TECHNIQUES

5. Some Basic Tools: How to Cut, Paste, Copy, Measure, and Visualize DNA

6. Cloning and Expression Vectors, Libraries, and Their Screening

Part III: Applications in Entomology

10. Sex Determination in Insects

11. Molecular Genetics of Insect Behavior

12. Molecular Systematics and Evolution of Arthropods

13. Insect Population Ecology and Molecular Genetics

PREFACE TO THE SECOND EDITION

Amazing progress has been made in insect molecular genetics since the first edition appeared in 1994. Transformation of insects other than *Drosophila melanogaster* has become an almost routine project. The *Drosophila* Genome Project was completed in 1999 and produced many surprises and promises a fruitful future for mining genes and developing an understanding of genome structure, function, and evolution. The mining of this treasure trove of data will require some years of work, but the possibility exists that we ultimately will be able to understand how this insect develops. Insect biology will become synthetic again with the use of genomics, transcriptomics, and proteomics approaches. The complete sequencing of other complex eukaryotic genomes, including those of *Caenorhabditis elegans* and *Homo sapiens*, opened additional doors to compare genome organization, evolution, and gene function.

Molecular methods and technology have changed rapidly in the past few years, with a plethora of new kits available for extracting and purifying DNA and RNA, for cloning, sequencing, and amplifying DNA and RNA by the polymerase chain reaction (PCR). Gene chip or microarray methods offer new tools for learning about gene function. All the improvements in these molecular toolkits make molecular methods ever more accessible to the entomological community.

The same basic organization with three major sections has been maintained in this edition, but the chapters have been updated with recent references. References were included that provide an entry into the recent literature; where possible, review articles are cited. I regret that I could not include references to all the new molecular studies on insects; there are just too many! That alone signals that molecular entomology is maturing.

This book is dedicated to entomologists just beginning their research careers; I hope this book helps you to start exciting and productive projects that employ these valuable molecular tools. For those of you with no background in molecular genetics, the book should be read from start to finish. Key concepts are highlighted in the "Overview" and reading it both prior to and after reading each chapter may be helpful. The diagrams, especially those illustrating molecular methods, should be evaluated while reading the text. In many cases, the concepts involved are most readily obtained if the text and diagrams are read together.

Finally, I thank all the people who so kindly provided feedback on the first edition and those who made thoughtful suggestions on earlier drafts of this one, including Anna Malacrida, David Haymer, A. Jeyaprakash, Lucy Skelley, Juan Alvarez, Jim Hoy, and Alison Walker. I sincerely thank those who kindly provided illustrations and Pam Howell,

Mike Sanford, and Pat Hope for their assistance in getting this manuscript and its illustrations completed. This is Florida Agricultural Experiment Station Journal Series R-08721.

Marjorie A. Hoy
Department of Entomology and Nematology
University of Florida, Gainesville

PREFACE TO THE FIRST EDITION

The development of recombinant DNA techniques during the past 20 years has resulted in exciting advances in the detailed study of specific genes at the molecular level as well as breakthroughs in molecular, cellular, and developmental biology. Of the molecular genetics studies conducted on insects, most have been directed to *Drosophila melanogaster*. Relatively few data have been generated by molecular biological methods from analyses of other insects. Yet, the application of molecular genetics to insects other than *Drosophila* has the potential to revolutionize insect population and organismal biology.

Why have molecular genetic techniques been used so little by entomologists? There may be a number of reasons. Recombinant DNA techniques are most readily carried out by people trained in biochemistry and relatively few entomologists are so trained. The techniques have been, until recently, relatively complex and difficult, so that strong technical skills were required. Also, most entomologists have been slow to ask whether these techniques were appropriate for studies of population or organismal biology because much of the published literature has focused on fundamental issues of *Drosophila* gene structure, regulation, and function, developmental regulation, and evolution.

Goals

My goal is to introduce entomologists to the concepts of molecular genetics without assuming that they have received previous training in molecular biology. This book is not intended to substitute for formal training in biochemistry or molecular genetics. If novice readers wish to develop molecular genetics skills, they must obtain additional training in genetics and biochemistry. However, the book will provide an introduction to terminology, as well as an overview of principles, techniques, and possible applications of molecular genetics to problems of interest to entomologists.

In preference to using examples from the *Drosophila* literature, I have used examples in which other arthropods have been studied. However, without doubt, *Drosophila* is the premier model for insect molecular genetics study. One fond hope is that this book will be a bridge for entomologists seeking to apply the exciting methods developed for *Drosophila* and that it will introduce *Drosophila* workers to some of the problems and issues of interest to entomologists seeking to solve applied problems. Perhaps this book will help to break down the barriers between entomologists and *Drosophila* workers isolated from each other

by perspective and technical jargon. If this book helps to achieve these goals, it will have served its purpose.

Organization

The book was designed for a one-semester course in insect molecular genetics for upper-division undergraduates or beginning graduate students. The initial portion of the book reviews basic information about DNA, RNA, and other important molecules (Chapters 1–4). Readers with a recent course in genetics could skip this section. Chapter 5 describes the genetic systems found in insects and an overview of development sufficient to understand subsequent techniques such as P-element-mediated transformation and sex determination. Chapters 6–9 provide introductions to useful techniques, including cloning, library construction, sequencing, the polymerase chain reaction, and P-element-mediated transformation of *Drosophila*. Most molecular biologists reading this book could skip this section as well. Chapters 6–9 are not intended as a laboratory manual but, in some cases, an outline of laboratory protocols is provided in order to furnish the novice with a sense of the complexity or simplicity of the procedures and some of the issues to consider in problem solving. Throughout the book, references are provided for the reader interested in pursuing specific topics and techniques, although they are not exhaustive. Despite the value of providing an historical overview, I have not always provided references to the first publication on a subject. Rather, review articles or recent publications that include references to earlier work are cited.

Finally, in the third section of the book (Chapters 10–14), I have attempted to demonstrate how molecular genetic techniques can solve a diverse array of basic and applied problems. Part III is intended to introduce readers to the exciting molecular research that is revolutionizing insect biology, ecology, systematics, behavior, physiology, development, sex determination, and pest management. Each chapter in this section could be read by itself, assuming that reader understands the appropriate concepts or information presented in Parts I and II.

Each chapter begins with an overview or brief summary of the material being covered. The overview should be read both before and after reading each chapter to review the concepts covered. The overview is followed by a brief introduction covering the history or rationale for the topic. References at the end of the chapter are provided for further reading. Where possible, books or reviews are cited to provide an entry into the literature. Recent references are provided, but no attempt has been made to review all the literature on a specific topic. Simple protocols may be given to provide the flavor of specific techniques, although these are not intended to be complete. References to handbooks or techniques books are also provided at the end of appropriate chapters. When a term that may be unfamiliar is first introduced, it written in boldface and a brief definition or description is given in the Glossary at the end of the book. Finally, in Appendix I, a time line of some significant advances in genetics, molecular biology, and insect molecular genetics provides a perspective of the pace with which dramatic advances have been, and continue to be, made.

Progress is rapid in molecular genetics, and this book can only provide an introduction to the principles of insect molecular genetics and some of its applications. It is impossible to provide a complete review of the insect molecular genetics literature in a book of this size. The literature cited includes references through 1993 and focuses on genetics. It is not intended to be an introduction to all aspects of "molecular entomology," which has been

defined as "a blend of insect science, molecular biology, and biochemistry." The dividing line between molecular entomology and insect molecular genetics is sometimes difficult to resolve.

Shortly before this book went to the publisher, two related books were published: *Molecular Approaches to Fundamental and Applied Entomology*, edited by J. Oakeshott and M. J. Whitten, and *Insect Molecular Science*, edited by J. M. Crampton and P. Eggleston. Both multiauthored books cover some of the topics included here, but assume the reader is familiar with molecular genetic techniques and terminology; they would be daunting for the novice.

1

DNA, Gene Structure, and DNA Replication

1.1. Overview

Arthropod genes are made of DNA and are located in chromosomes that consist of proteins, RNA, and DNA. DNA is a polymer of nucleotides. Each nucleotide consists of a pentose sugar, one of four nitrogenous bases, and a phosphoric acid component. DNA consists of two complementary strands in a helix form. Pairing of the nitrogenous bases adenine (A) with thymine (T) and cytosine (C) with guanine (G) on the two complementary strands occurs by hydrogen bonding. A pairs with T by two hydrogen bonds, and C pairs with G by three hydrogen bonds. DNA has chemically distinct 5′ and 3′ ends. The two strands are antiparallel, with one running in the 5′ to 3′ direction and the other from the 3′ to 5′ direction. The antiparallel orientation of the two strands creates a special problem when the DNA is duplicated or replicated during mitosis or meiosis.

Genetic information is determined by the sequence of nitrogenous bases (A, T, G, C) in one of the strands, with a three-base (triplet) codon designating an amino acid. The genetic code is degenerate, which means that more than one codon specifies most amino acids. The genetic information is expressed when DNA is transcribed into messenger RNA, which then is translated into polypeptides. Most insect genes have intervening noncoding sequences (introns) that must be removed from the primary RNA molecule before translation into the protein can occur.

Efficient and accurate replication of DNA must occur at each cell division, or the cell or organism may not survive. DNA replication is semiconservative, which means that one of the nucleotide strands of each new DNA molecule is new and the other is old in each "cell generation." The new DNA strand is complementary to the parental (or template) strand. DNA replication occurs in one direction only, from the 5′ to the 3′ end of the strand, and thus replication takes place differently on the two antiparallel strands. Replication on the "leading strand" can occur in the 5′ to 3′ direction in a continuous manner. However, DNA replication on the other strand, the "lagging strand," occurs in short segments (Okazaki fragments) because the DNA runs in the 3′ to 5′ direction. Subsequently, the Okazaki fragments must be ligated together. Replication of DNA in chromosomes begins at multiple sites, called origins of replication, along the chromosome and involves a number of enzymes and proteins. Although DNA replication is usually highly accurate, errors in DNA replication, or mutations, can result from duplications, deletions, inversions, and translocations of nucleotides, which may affect the functioning of the resultant polypeptide. New combinations of genes can occur through recombination during meiosis.

1.2. Introduction to the Central Dogma

The **Central Dogma**, as proposed by Francis Crick in 1958, stated that biological information is carried in DNA, and that this information subsequently is transferred to RNA and finally to proteins. Initially, the Central Dogma stated that the flow of information is unidirectional, with proteins unable to direct synthesis of RNA, and RNA unable to direct the synthesis of DNA (Figure 1.1).

The Central Dogma had to be amended in 1970 when certain viruses were found to transfer information from RNA to DNA. Subsequently, mutated proteins in the membrane of brain cells of vertebrates were found to be "inherited." Although such aberrant proteins initially were thought to be caused by "slow viruses" or viroids, Stanley Prusiner discovered that those mutated proteins (called **prions**) could cause a group of invariably

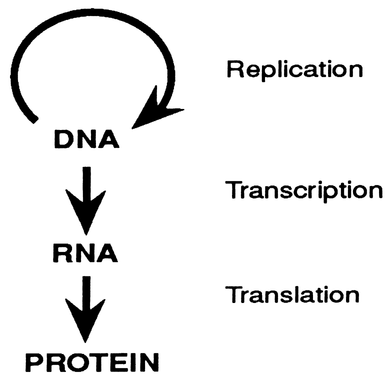


Figure 1.1. The Central Dogma assumes that biological information is transfered is from DNA to RNA to proteins. Recent discoveries of viruses that transcribe information from RNA to DNA has required modification in the Dogma. Three processes are involved in the Central Dogma: DNA replication, transcription of the genetic information into RNA, and translation of the messenger RNA into a polypeptide (protein).

fatal neurodegenerative diseases. The term *prion* refers to proteinaceous infectious particles (Prusiner and Scott 1997). Prion diseases include bovine spongiform encephalopathy (BSE or “mad cow disease”) in cattle, scrapie in sheep, and Creutzfeld–Jakob Disease or kuru in humans. These “proteinaceous infective particles” do *not* contain DNA, but are able to transmit the disease to other individuals who eat the altered proteins (Prusiner and Scott 1997). Current data suggest the altered protein acts as a template upon which the normal protein is refolded into a deformed molecule through a process facilitated by another protein (Prusiner and Scott 1997, Tuite 2000). Such abnormal proteins are transmitted to daughter cells, thus propagating the mutant phenotype in the absence of mutated nucleic acid.

Despite these exceptions, the Central Dogma remains a major tenet of modern biology. In insects, the genes (DNA) are found in complex structures called chromosomes that consist of proteins, RNA, and DNA. This chapter reviews the structure of DNA and RNA, the basis of the genetic code, the processes involved in DNA replication, and changes in DNA that result in mutations.

1.3. The “RNA World” Came First?

It is now widely accepted that there was an era on Earth during which RNA played the role of both genetic material and main agent of catalytic activity, e.g., had ribozyme activity (DiGiulio 1997, Jeffares et al. 1998, Poole et al. 1998, Cooper 2000, Eddy 2001). This implies that proteins in the modern world replaced RNA as the main catalysts (enzymes). The “RNA organism” is thought to have had a multiple-copy, double-stranded RNA genome capable of recombination and splicing. The RNA genome was probably fragmented into “chromosomes” (Jeffares et al. 1998). RNA could have been the first genetic material because we now know it can serve as a template for self-replication and can catalyze a number of chemical reactions, including the polymerization of nucleotides (Johnston et al. 2001). It is thought that interactions between RNA and amino acids then evolved into the present-day world in which DNA is the primary stable repository of genetic information.

1.4. The Molecular Structure of DNA

Deoxyribonucleic acid (DNA) is a long polymeric molecule consisting of numerous individual monomers that are linked in a series and organized in a helix. Each monomer is called a **nucleotide**. Each nucleotide is itself a complex molecule made up of three components: (1) a sugar, (2) a nitrogenous base, and (3) a phosphoric acid.

In DNA, the sugar component is a pentose (with five carbon atoms) in a ring form that is called 2′-deoxyribose (Figure 1.2).

The nitrogenous bases are single- or double-ring structures that are attached to the 1′-carbon of the sugar. The bases are **purines** (adenine and guanine) or **pyrimidines** (thymine and cytosine) (Figure 1.3). When a sugar is joined to a base, it is called a **nucleoside**.

A nucleoside is converted to a nucleotide by the attachment of a **phosphoric acid group** to the 5′-carbon of the sugar ring (Figure 1.4). The four different nucleotides that polymerize to form DNA are 2′-deoxyadenosine 5′-triphosphate (dATP or A), 2′-deoxyguanosine 5′-triphosphate (dGTP or G), 2′-deoxycytidine 5′-triphosphate (dCTP or C), and 2′-deoxythymidine 5′-triphosphate (dTTP or T) (Figure 1.5). These names are usually abbreviated as dATP, dGTP, dCTP, and dTTP, or shortened further as A, G, C, and T.

Individual nucleotides are linked together to form a polynucleotide by **phosphodiester bonds** (Figure 1.4). Polynucleotides have chemically distinct ends. In Figure 1.5, the top of the polynucleotide ends with a nucleotide in which the triphosphate group attached to the 5′-carbon has not participated in a phosphodiester bond. This is called the 5′ or 5′-P terminus. At the other end of the molecule the unreacted group is not the phosphate, but the 3′-hydroxyl. This is called the 3′ or 3′-OH terminus. This distinction between the two ends (5′ and 3′) means that polynucleotides have an orientation that is very important in many molecular genetics applications.

Polynucleotides can be of any length and have any sequence of bases. The DNA molecules in chromosomes are probably several million nucleotides long. Because there are no restrictions on the nucleotide sequence, a polynucleotide just 10 nucleotides long could have any one of 4^{10} (or 1,048,576) different sequences. This ability to vary the sequence is what allows DNA to contain complex genetic information.

1.5. The Molecular Structure of RNA

RNA is also a polynucleotide, but with two important differences from the structure of DNA. First, the sugar in RNA is **ribose** (Figure 1.2). Second, RNA contains the nitrogenous base

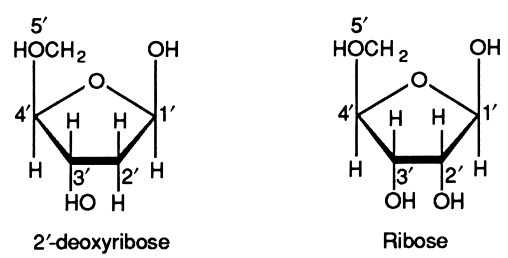


Figure 1.2. Structure of sugars found in nucleic acids; 2′-deoxyribose is found in DNA and ribose is found in RNA.

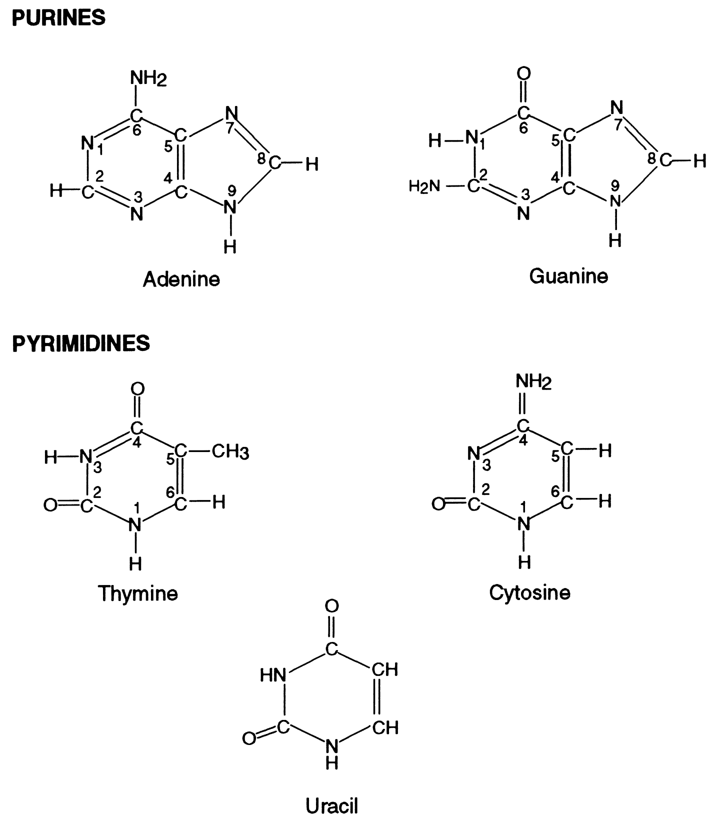


Figure 1.3. Bases in DNA are purines (adenine and guanine) or pyrimidines (thymine and cytosine). Uracil is substituted for thymine in RNA.

uracil (U) instead of thymine (Figure 1.3). The four nucleotides that polymerize to form RNA are adenosine 5′-triphosphate, guanosine 5′-triphosphate, cytidine 5′-triphosphate, and uridine 5′-triphosphate, which are abbreviated as ATP, GTP, CTP, and UTP or A, G, C, or U. The individual nucleotides are linked together with 3′ to 5′ phosphodiester bonds. RNA is typically single-stranded, although it can form complex structures (such as hairpins) or become double-stranded under some circumstances.

1.6. The Double Helix

The discovery, by Watson and Crick (1953), that DNA is a double helix of antiparallel polynucleotides ranks as one of the most important discoveries in biology. Nitrogenous bases are located inside the double helix, with the sugar and phosphate groups forming the backbone of the molecule on the outside (Figure 1.6). The nitrogenous bases of the two polynucleotides interact by **hydrogen bonding**, with an adenine (A) pairing to a thymine (T) and a guanine (G) to a cytosine (C).

Hydrogen bonds are weak bonds in which two negatively charged atoms share a hydrogen atom between them. Two hydrogen bonds form between A and T, and three between G and C.

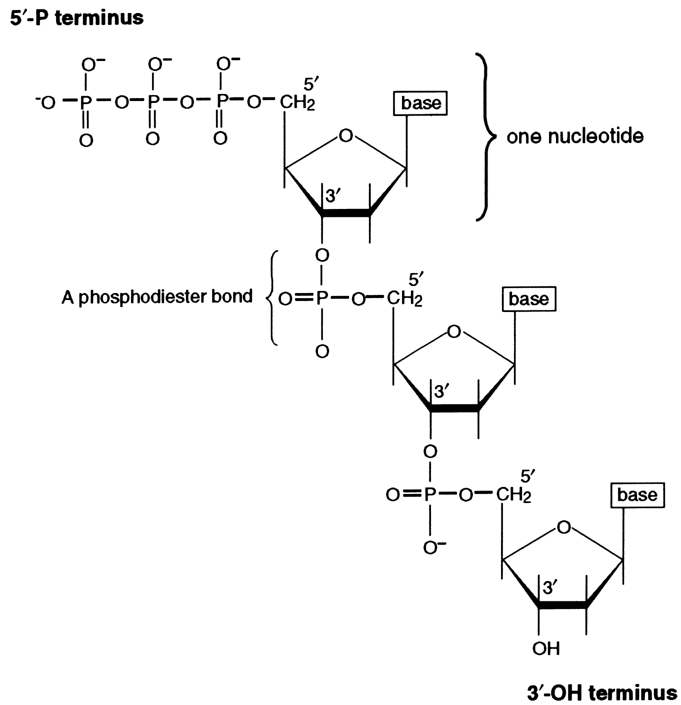


Figure 1.4. A nucleoside consists of a sugar joined to a base. It becomes a nucleotide when a phosphoric acid group is attached to the 5′-carbon of the sugar. Nucleotides link together by phosphodiester bonds to form polynucleotides.

Bonding between G and C is thus stronger, and more energy is required to break it. The hydrogen bonds, and other molecular interactions called stacking interactions, hold the double helix together.

The DNA helix turns approximately every 10 base pairs (abbreviated as 10 bp), with spacing between adjacent bp of 3.4 angstroms (Å) so that a complete turn requires 34 Å (Figure 1.6). The helix is 20 Å in diameter and right handed. This means that each chain follows a clockwise path. The strands run antiparallel to each other, with one running in the 5′ to 3′ direction and the other in the 3′ to 5′ direction. The DNA helix has two grooves, a **major** and a **minor groove** (Figure 1.6). Proteins involved in DNA replication and transcription often interact with the DNA and each other within these grooves.

1.7. Complementary Base Pairing Is Fundamental

The principle of **complementary base pairing** is a fundamental element of DNA and of great practical significance in many techniques used in genetic engineering. A pairs with T and G pairs with C. Normally, no other base pairing pattern will fit in the helix or allow hydrogen bonding to occur (Figure 1.7).

Complementary base pairing provides the mechanism by which the sequence of a DNA molecule is retained during replication of the DNA molecule, which is crucial

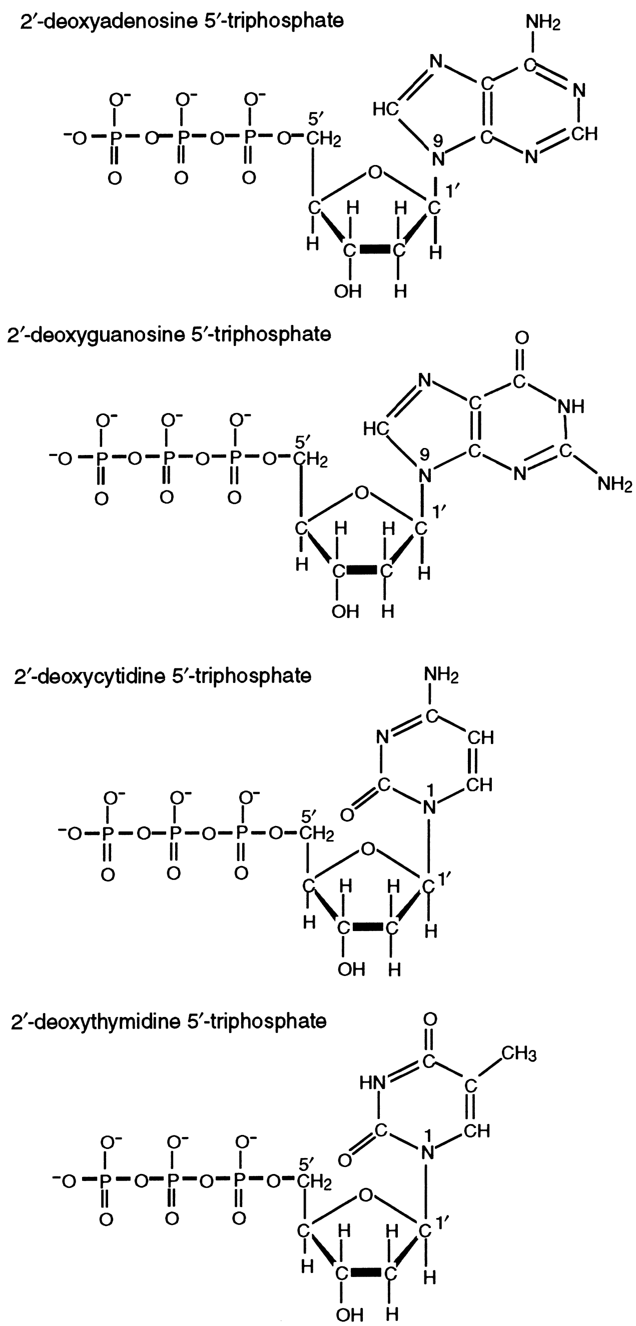


Figure 1.5. The four trinucleotides from which DNA is synthesized are 2′-deoxyadenosine 5′-triphosphate (dATP), 2′-deoxyguanosine 5′-triphosphate (dGTP), 2′-deoxycytidine 5′-triphosphate (dCTP), and 2′-deoxythymidine 5′-triphosphate (dTTP).

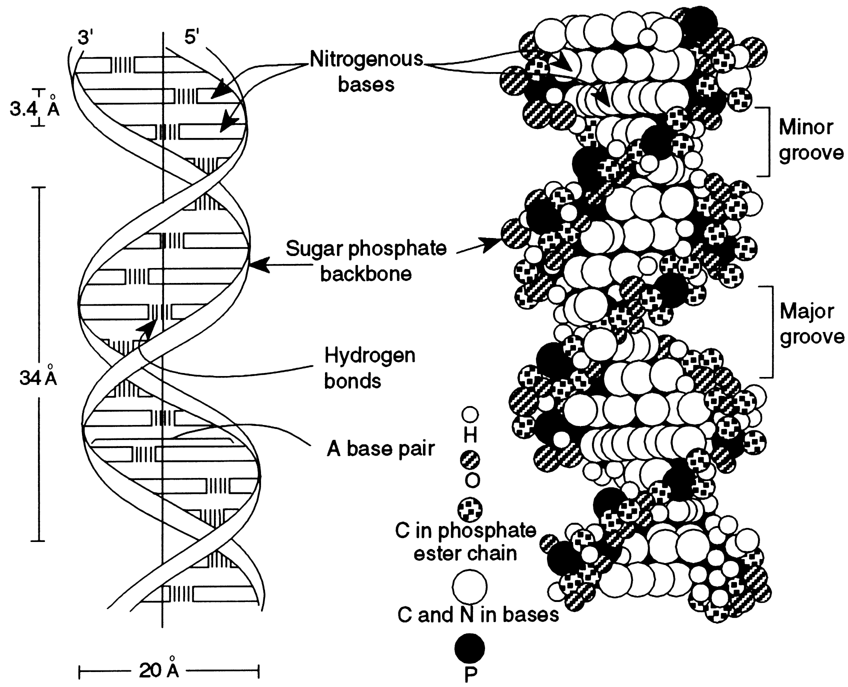


Figure 1.6. Two representations of the double helix structure of DNA. The model on the left shows the hydrogen bonding between nitrogenous bases that holds the two antiparallel strands together. The model on the right shows the relative sizes of the atoms in the molecule.

if the information contained in the gene is not to be altered or lost during cell division. Complementary base pairing is also important in the transcription and expression of genetic information in the living insect.

1.8. DNA Exists in Several Forms

DNA actually is a dynamic molecule in living organisms and has several different variations in form. In some regions of the chromosome, the strands of the DNA molecule may separate and later come back together. DNA typically is right-handed, and it can form more than 20 slightly different variations of right-handed helices. In some regions of the molecule, it can even form left-handed helices. If segments of nucleotides in the same strand are complementary, the DNA may even fold back upon itself in a hairpin structure.

DNA exists in different crystalline forms, depending upon the amount of water present in the DNA solution. The B form is the structure in which DNA commonly occurs under most cellular conditions. A-DNA is more compact than B-DNA, with 11 bp per turn of the helix and a diameter of 12 Å. In addition, C-, D-, E-, and Z-DNA have been found. The Z-DNA form has a left-handed helix rather than a right-handed helix. A triple helical form (H) also occurs. A, H, and Z forms are thought to occur in cells, and C, D, and E forms of DNA may be produced only under laboratory conditions.

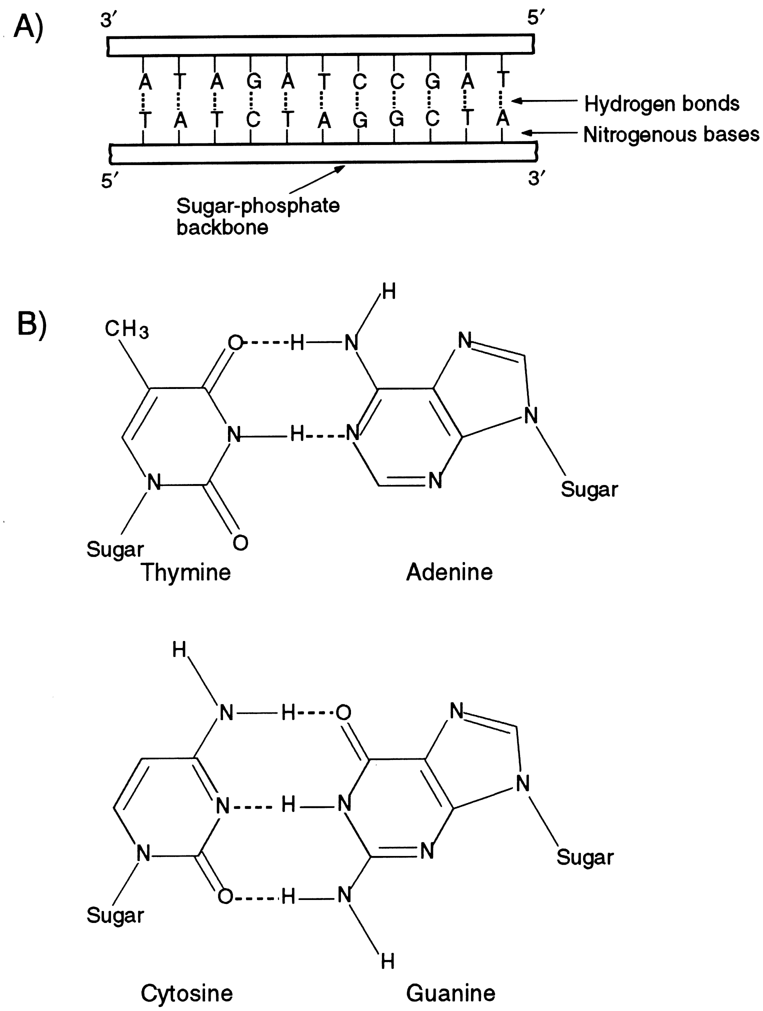


Figure 1.7. A) Complementary base-pairing of polynucleotides by hydrogen bonds holds the two strands of the DNA molecule together. B) Thymine (T) pairs with adenine (A) with two hydrogen bonds, and guanine (G) pairs with cytosine (C) with three.

1.9. Genes

The concept of a "gene" has evolved as genetics has changed (Muller 1947, Maienschein 1992). Until 1944, when Avery et al. (1944) demonstrated that the genetic information resided in nucleic acids, it was considered possible that the genetic information was encoded in proteins. "Genes" can be a specific location on a chromosome, a particular type of biochemical material, and a physiological unit that directs development. Genes are segments of a DNA molecule, which may vary in size from as few as 75 nucleotides (nt) to more than 200 **kilobases** (kb) of DNA. (A kilobase is 1000 nucleotides.) Genes contain biological information by coding for the synthesis of an RNA molecule. The RNA may subsequently direct the synthesis of an enzyme or other protein molecule. RNA also may be used directly as the

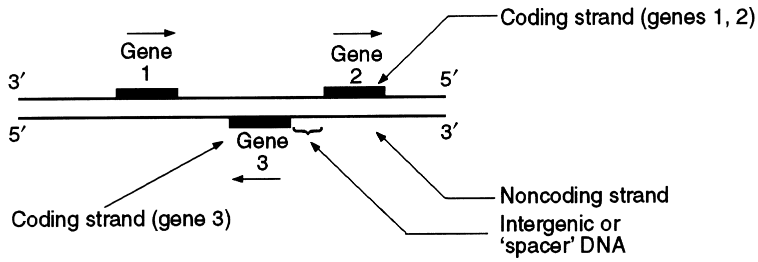


Figure 1.8. Genetic information is contained in genes carried on one of the two strands (coding strand). The complementary strand *in that region* is the noncoding strand. Genes can occur on different strands at different points of the DNA molecule. Noncoding DNA between genes is called intergenic or spacer DNA.

gene product itself, e.g., as transfer RNA, ribosomal RNA, small nucleolar RNA, and small nuclear RNA (Eddy 2001). Proteins may regulate other genes, form part of the structure of cells, or function as enzymes. Expression of the information contained in protein-coding genes involves a two-step process of **transcription** and **translation** (Figure 1.1).

We now know that the actual genetic information is determined by just one of the two polynucleotide strands of the double-helix DNA molecule. This is called the **coding strand**, and the other strand is the noncoding complement to it. Sometimes the coding strand is known as the **sense** strand and the noncoding as the **antisense** strand. A few examples are known in which both strands in a specific region code for different genes. Often one strand of the double helix may be the sense strand over part of its length but be the antisense strand over other segments (Figure 1.8). As you can see, the definition of a "gene" is complex and has changed through time (Eddy 2001, Nelkin 2001). A protein-coding gene typically includes a variety of regulatory structures and signals, as will be described in Chapter 2.

1.10. The Genetic Code Is a Triplet and Is Degenerate

The genetic code for a protein-coding gene is based on the sequence of three nucleotides in the DNA molecule. The triplet sequence (or **codon**) determines which amino acids are assembled in a particular sequence into proteins. It is possible to order four different bases (A, T, C, G) in combinations of three into 64 triplets or codons. However, there are only approximately 20 different amino acids, so the question immediately arises: what *do* the other 44 codons do?

The answer is that the code is degenerate, with all amino acids except methionine and tryptophan determined by more than one codon (Table 1.1). The codons in Table 1.1 are represented by A, U, C, and G because the genetic information in DNA is transcribed by messenger RNA, which uses U instead of T.

The genetic code contains punctuation codons. Three different codons (UAA, UGA, and UAG) function as "stop" messages or **termination codons**; they occur at the end of a protein-coding gene to indicate where translation should stop. AUG serves as an **initiation** or **start codon** when it occurs at the front end of a gene. Because AUG is the sole codon for the amino acid methionine, AUGs also are found in the middle of genes.

Table 1.1. The 20 Amino Acids That Occur in Proteins and Their Codons

Amino acid	Abbreviations		Codons					
Alanine	ala	A	GCU	GCC	GCA	GCG		
Arginine	arg	R	AGA	AGG	CGU	CGC	CGA	CGG
Asparagine	asn	N	AAU	AAC				
Aspartic acid	asp	D	GAU	GAC				
Cysteine	cys	C	UGU	UGC				
Glutamic acid	glu	E	GAA	GAG				
Glutamine	gln	Q	CAA	CAG				
Glycine	gly	G	GGU	GGC	GGA	GGG		
Histidine	his	H	CAU	CAC				
Isoleucine	ile	I	AUU	AUC	AUA			
Leucine	leu	L	UUA	UUG	CUU	CUC	CUA	CUG
Lysine	lys	K	AAA	AAG				
Methionine[a]	met	M	AUG					
Phenylalanine	phe	F	UUU	UUC				
Proline	pro	P	CCU	CCC	CCA	CCG		
Serine	ser	S	AGU	AGC	UCU	UCC	UCA	UCG
Threonine	thr	T	ACU	ACC	ACA	ACG		
Tryptophan[a]	trp	W	UGG					
Tyrosine	tyr	Y	UAU	UAC				
Valine	val	V	GUU	GUC	GUA	GUG		

[a] Methionine and tryptophan are underlined because they are specified by only one codon.

The genetic code is not universal, although it was assumed to be so initially. In 1979 it was found that mitochondrial genes use a slightly different code (Knight et al. 2001). For example, the codon AGA typically codes for arginine, but in *Drosophila* mitochondria the codon AGA codes for serine.

The thesis of three primary domains of life has revolutionized theories of cell evolution. The three-domain concept of life includes the Archaea (archaebacteria), Bacteria (eubacteria), and Eukarya (eukaryotes). **Eukaryotes** are organisms (including insects) that consist of cells with true nuclei bounded by nuclear membranes. Cell division in eukaryotes occurs by mitosis, reproductive cells undergo meiosis, and oxidative enzymes are packaged in mitochondria. Evidence continues to grow indicating that eukaryotic genes are derived from both the archaebacterial (informational genes) and the eubacterial (operational genes) lineages, indicating that eukaryotic genomes are chimeric (Lang et al. 1999, Nesbo et al. 2001).

1.11. Gene Organization

Genes are located on chromosomes. Each chromosome contains a single DNA molecule. These DNA molecules contain hundreds or thousands of genes. The fruit fly *Drosophila melanogaster* is estimated to have approximately 13,600 distributed on four chromosomes (Adams et al. 2000). Genes may be spaced out along the length of a DNA molecule with noncoding DNA sequences intervening, or the genes may be grouped into clusters. Genes in a cluster may be related or unrelated to each other in structure and function.

There usually are segments of DNA in eukaryotes in which the nucleotide sequences apparently do not code for anything; this DNA is called “spacer” DNA if it occurs between genes.

Multigene families are clusters of related genes with similar nucleotide sequences. Multigene families may have originated from a single ancestral gene that duplicated to produce two, or more, identical genes. These identical genes could have diverged in nucleotide sequence through time to produce (two or more) related functional genes. In many cases, the genes of multigene families have become scattered at different positions on more than one chromosome by large-scale rearrangements (translocations or inversions) that occur both within and between chromosomes. Examples of multigene families in insects include *actins*, *tubulins*, *heat shock*, salivary glue, *chorion*, cuticle, and yolk protein genes. (Note that the name of a specific gene usually is italicized.)

Pseudogenes are DNA sequences that appear similar to those of functional genes, but the genetic information has been altered so that the former gene is no longer functional. Once the biological information has been lost, a pseudogene can undergo rapid changes in nucleotide sequence and, given sufficient time, may evolve to the point where it is not possible to identify it as a former gene. At this point it might be called "junk" DNA.

One of the more interesting recent discoveries in genetics was the revelation in 1977 that most protein-coding genes in eukaryotic organisms are discontinuous. **Discontinuous genes** contain coding and noncoding segments called **exons** and **introns**, respectively (Figure 1.9).

Considerable discussion of the origin, evolution and importance of introns has occurred (Herbert 1996, Gilbert et al. 1997, Trotman, 1998). Introns have been maligned as examples of "junk" DNA because they may be considerably longer than the coding sequences (exons) and have no apparent function. Two major hypotheses have been proposed to explain the origin of introns: one is the "introns-early" hypothesis.

Under the **introns-early** hypothesis, many introns were present in the common ancestor of all life, but large or complete losses of the introns occurred in many independent lineages. According to this hypothesis, introns functioned in the primordial assembly of protein genes by promoting the recombination, or shuffling, of short exons, each encoding 15 to 20 amino acids (minigenes) into different functional genes through fusion (Gilbert et al. 1997). It is likely that there has been an average of two or three acts of such fusions of minigenes into the larger exons of today (Gilbert et al. 1997).

Some introns have been inherited for millions of years, making it possible to find a consistent location for the introns when homologous genes from different organisms are examined. The actual sequences of the introns in these homologous genes may have diverged through mutation to the point that they appear to have no sequence similarity. Trotman (1998) suggests that this consistent location of introns is evidence that introns may have been integral to the development of primordial genes. However, very few of the ancient introns may have survived in a detectable form because cellular life is very old and evolution has had a long time to change the sequences within these noncoding regions (Trotman 1998).

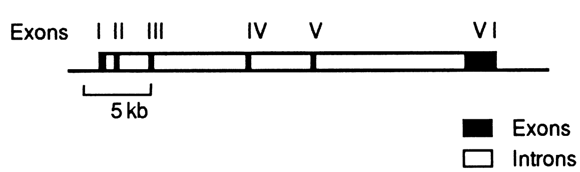


Figure 1.9. Protein-coding genes in eukaryotic organisms are divided into introns and exons. Introns are removed from the messenger RNA before it is translated into a polypeptide. In this example, there are six exons and five introns. The genetic message is present in exons I, II, III, IV, V, and VI.

The **introns-late** hypothesis assumes that mechanisms for splicing introns out were not present in the common ancestor of life, but arose and spread within eukaryotes during their evolution. Under this hypothesis, introns could not have played a role in ancient gene and protein assembly.

As is often the case with many "either/or" debates, the truth may be a combination of the two hypotheses. Tyshenko and Walker (1997) suggested both concepts may be correct; the introns in the *triosephosphate isomerase* genes of insects may be the result of transposable element insertion relatively recently, whereas other introns may have been present for a very long time (Logsdon et al. 1995). DeSouza et al. (1998) suggest that 30 to 40% of the present-day intron positions in ancient genes correspond to the introns originally present in the ancestral gene. The rest of the intron positions are due to the movement or addition of introns over evolutionary time. Thus, introns may be both early and late, with about 65% of the introns having been added to preexisting genes.

Introns generally are absent in the genes of prokaryotes and are rare in some eukaryotes, such as yeast. The number of introns and their lengths vary from species to species and from gene to gene. Some genes in eukaryotic organisms lack introns, whereas other genes in the same species may have as many as 50. Introns may interrupt a coding region, or they may occur in the untranslated regions of the gene. Some eukaryotic genes contain numerous and very large introns, but introns typically range from 100 to 10,000 bp in length. Oddly, a few introns contain genes themselves; how the genes got into the middle of an intron of another gene remains a mystery.

The presence of introns within many eukaryotic protein-coding genes requires that an additional step takes place between transcription and translation in eukaryotes. Thus, when the DNA is transcribed into RNA, the initial RNA transcript is not **messenger RNA** (**mRNA**). It is a precursor to mRNA and must undergo processing (splicing) in the nucleus to remove the introns before it appears in the cytoplasm as mRNA. This process is described in Chapter 2, but first DNA replication is reviewed.

1.12. Efficient DNA Replication Is Essential

Every living organism must make a copy of its genes in each cell each time the cell divides. Such replication ideally will be both rapid and accurate. If not, the organism′s survival and integrity are jeopardized. Even a very small error rate of 0.001% (one mistake per 100,000 nucleotides) can lead to detrimental changes. However, while many changes in DNA (called mutations) are detrimental, many apparently are neutral, and a few are beneficial.

Until recently, most studies of DNA replication were conducted using prokaryotic organisms such as the bacterium *Escherichia coli*, because *E. coli* is more easily studied. DNA replication in *E. coli*, which has a single circular DNA molecule, illustrates the basic principles. The complications added by arranging DNA molecules into linear chromosomes will then be described, although gaps in knowledge of eukaryotic DNA replication remain (Cook 1999, Berezney et al. 2000, Bogan et al. 2000, Sutton, 2001).

1.13. DNA Replication Is Semiconservative

DNA replication is **semiconservative**. This means the daughter molecules each contain one polynucleotide derived from the original DNA molecule and one newly synthesized strand

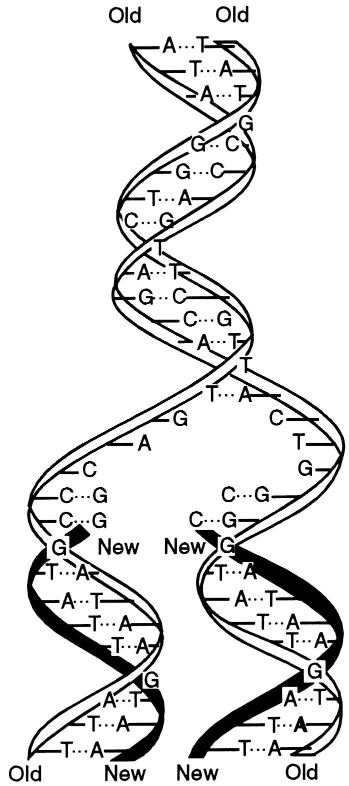


Figure 1.10. DNA replication is semiconservative, meaning that each new DNA helix contains one old and one new complementary strand. DNA synthesis relies on complementary base pairing to replicate DNA accurately.

(Figure 1.10). Semiconservative DNA replication requires the base pairing that holds the two strands together be broken so that synthesis of new complementary strands can occur.

1.14. Replication Begins at Replication Origins

During the replication of long DNA molecules, only a limited region of the DNA molecule is in an unpaired form at any one time. Replication occurs after the two strands separate, which involves breaking the weak hydrogen bonds holding the bases of the opposite strands together. The separation of the two strands starts at specific multiple positions in the chromosome called **origins of replication** and moves along the molecule. Synthesis of the new complementary polynucleotides occurs as the double helix "unzips." The region at which the base pairs of the parent molecule are broken and the new polynucleotides are synthesized is the **replication fork** (Figure 1.11).

The base pairing of the two strands of the parent DNA molecule is broken by enzymes called **helicases**. Once the helicase has broken the hydrogen bonds holding the two strands together, **single-strand binding proteins (SSBs)** attach to the single-stranded DNA to prevent the two complementary DNA strands from immediately reannealing (Figure 1.11).

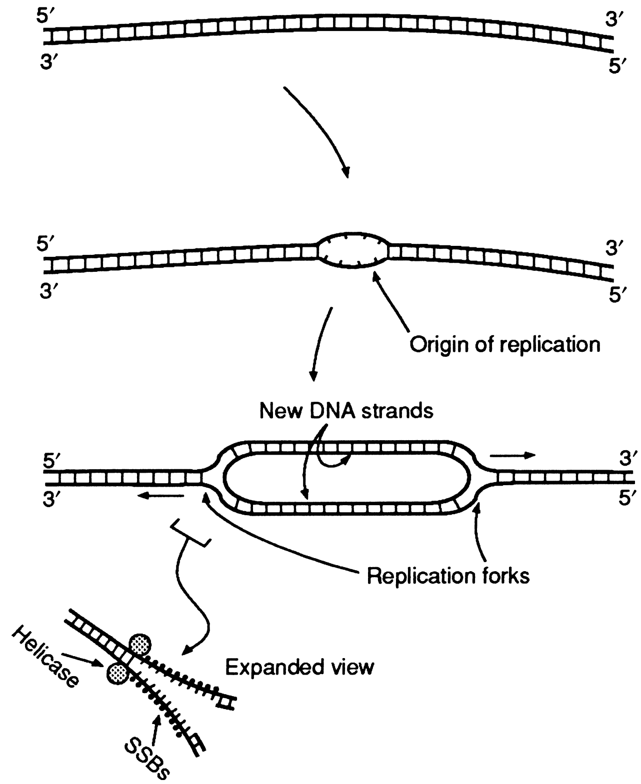


Figure 1.11. During DNA replication only part of the DNA molecule "unzips" to allow synthesis of new DNA strands. In this example, replication begins at an origin of replication. Eukaryotes have many origins of replication along their chromosomes. To keep the strands from reannealing at the replication forks where synthesis is occurring, single-strand binding (SSB) proteins attach. Helicases break the hydrogen bonds.

This makes it possible for **DNA polymerase** to synthesize new complementary DNA strands. We now know there are three distinct DNA polymerases in eukaryotes; one is important in initiating and priming replication, and two are involved in the main replication of DNA (Sutton and Walker 2001). One of these may function in replication of the leading strand, while the other may act on the lagging strand.

DNA polymerases have two properties that complicate DNA synthesis. First, DNA polymerase can synthesize *only* in the $5'$ to $3'$ direction, and, second, DNA polymerase cannot initiate the synthesis of new DNA strands without a **primer**.

1.15. Replication Occurs Only in the $5'$ to $3'$ Direction

Because DNA polymerases can synthesize DNA only in the $5'$ to $3'$ direction, the template strands must be read in the $3'$ to $5'$ direction. This is a straightforward process for one of the DNA template strands, called the **leading strand**, and DNA synthesis can proceed in an uninterrupted manner the entire length of the leading strand. However, DNA synthesis

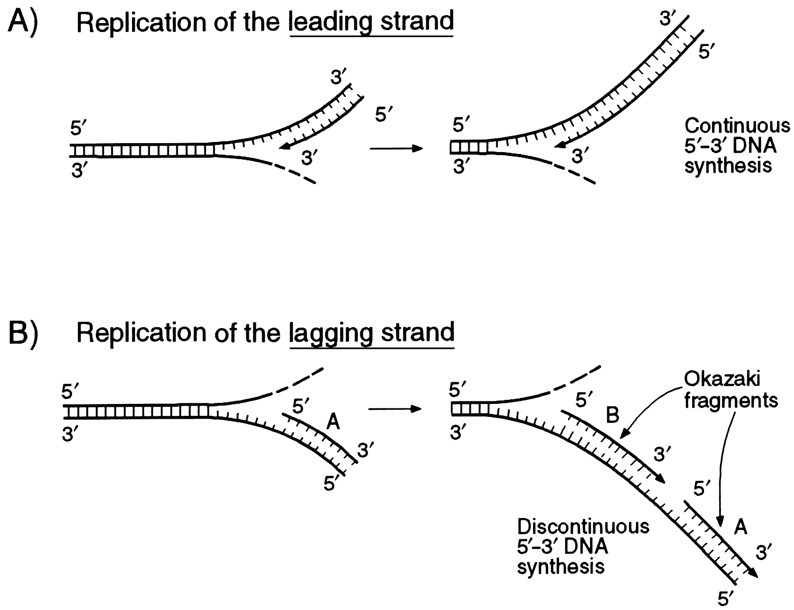


Figure 1.12. DNA replication occurs in a different manner on the two strands. A) The leading strand is continuously copied, with synthesis occurring in the 5′ to 3′ direction. B) Synthesis on the lagging strand is discontinuous. Synthesis occurs in short segments (Okazaki fragments) because DNA polymerase can only synthesize DNA in the 5′ to 3′ direction. Later these fragments are annealed together.

cannot proceed uninterrupted on the other template strand, called the **lagging strand** (Figure 1.12). DNA synthesis on the lagging strand is discontinuous, occurring in short sections, and produces short fragments of DNA called **Okazaki fragments**, after their discoverer who identified them in 1968.

1.16. Replication of DNA Requires an RNA Primer

Another complication of DNA synthesis is that synthesis is not initiated by DNA polymerase unless there is a short double-stranded region that can act as a **primer** (Figure 1.13). Apparently, the first few (50 to 75) nucleotides attached to either the leading or lagging strands are not deoxyribonucleotides, but rather ribonucleotides that are put in place by an RNA polymerase called **primase**. Once these ribonucleotides have been polymerized on the DNA template, the primase detaches, and polymerization of DNA is continued by DNA polymerase (Figure 1.13).

1.17. Ligation of Replicated DNA Fragments

After the Okazaki fragments (sequences complementary to the lagging strand of DNA) are produced, they must be joined together to produce a continuous strand (Figure 1.12). On the lagging strand, DNA polymerase III of *E. coli* stops when it reaches the RNA primer at

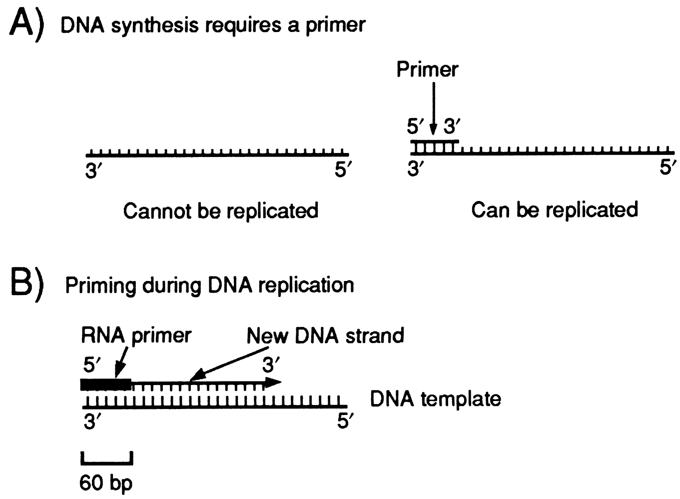


Figure 1.13. A) DNA must be primed or DNA polymerase is unable to synthesize a complementary strand. B) A primer of ribonucleotides is attached to a strand by RNA polymerase. DNA polymerase can then attach deoxyribonucleotides (dNTPs) to the DNA template in a sequence that is determined by the template strand. DNA synthesis occurs in the 5′ to 3′ direction.

the 5′ end of the next Okazaki fragment. Then DNA polymerase I of *E. coli* removes the ribonucleotides from the Okazaki fragment and replaces them with deoxyribonucleotides. When all the ribonucleotides have been replaced, DNA polymerase I replaces nucleotides on a short distance into the DNA region, before it dissociates from the new double-helix molecule. The Okazaki fragments then are joined up by **DNA ligase** which catalyzes the formation of a phosphodiester bond between the neighboring nucleotides.

DNA replication also requires that the double helix be unwound, as well as unzipped. There are approximately 400,000 turns in 400 kb of DNA. This unwinding is accomplished with the aid of enzymes called **DNA topoisomerases**. DNA topoisomerases unwind a DNA molecule without rotating the helix by causing short-term breaks in the polynucleotide backbone just in front of the replication fork. The reverse reaction is performed by DNA topoisomerases so that DNA molecules can be coiled.

1.18. DNA Replication in Eukaryotes

The replication of prokaryotic and eukaryotic DNA is similar, but differs in several aspects, the details of which are still being resolved (Gavin et al. 1995, Huberman 1995, Baker and Bell 1998, Leipe et al. 1999, Sutton and Walker 2001). DNA replication takes place during the eukaryotic cell cycle before the metaphase chromosomes become visible in mitosis or meiosis.

The **cell cycle** consists of three distinct phases. There are two gap periods (G_1 and G_2), when the cell is carrying out its normal metabolic activities (Figure 1.14), separated by the **S phase**, which is when DNA replication or synthesis occurs. Mitosis (M) occurs subsequent to the G_2 phase.

In order to reduce the amount of time required to replicate the very long DNA molecule in eukaryotic chromosomes, DNA replication is initiated at a series of **replication origins**

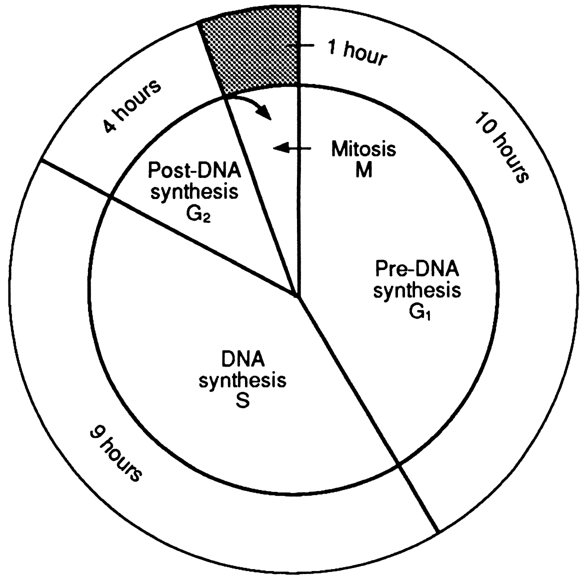


Figure 1.14. The cell cycle of a eukaryotic cell with a generation time of 24 hours. DNA synthesis occurs during the S phase. During G_1 and G_2 no DNA synthesis occurs. Mitosis (M) occurs after G_2.

about 40 kb apart on the linear chromosome and proceeds in both directions (Figure 1.14) (DePamphilis 1999). For example, replication in *Drosophila melanogaster* occurs at a rate of about 2600 nucleotide pairs per minute at 24°C. The largest chromosome in *Drosophila* is about 8×10^7 nucleotides long so, with about 8500 replication origins per chromosome, approximately 0.25 to 0.5 hour is required to replicate this chromosome. If replication occurred from a single replication fork, rather than from multiple replication origins, replication of a single chromosome would require about 15 days.

Several origins of replication (autonomously replicating sequences, or ARSs) of eukaryotic chromosomes have been identified in yeast. Functional ARS elements span about 100 base pairs, and many have an 11-base-pair core sequence that is essential for ARS function. The core sequence is the binding site of a protein complex (origin replication complex or ORC) that is essential for initiation of DNA replication at yeast origins. The ORC complex appears to recruit other proteins (including DNA helicases) to the origin of replication, leading to the start of replication. Proteins related to the yeast ORC proteins have been identified in *Drosophila* and other eukaryotes (Gavin et al. 1995).

Eukaryotes contain several different DNA polymerases. One is located in the mitochondria and is responsible for replication of mitochondrial DNA. The other DNA polymerases are in the nucleus and are involved in DNA replication. Polymerase α complexes with **primase** (the RNA polymerase that primes DNA synthesis) and appears to function with primase to synthesize short RNA–DNA fragments. Two other polymerases then synthesize the leading and lagging strands, extending the RNA–DNA primers initially synthesized by the polymerase α/primase complex. A DNA polymerase fills the gaps between the Okazaki fragments after the primers are removed (Sutton and Walker 2001).

Proteins (called sliding-clamp proteins and clamp-loading proteins) act at the eukaryote replication fork to load the polymerase onto the primer and maintain its stable association

with the template. The **clamp-loading proteins** (called replication factor C, or RFC) in eukaryotes recognize and bind DNA at the junction between the primer and template. The sliding-clamp proteins (proliferating cell nuclear antigen, or PCNA) in eukaryotes bind adjacent to the clamp-loading proteins, forming a ring around the template DNA. The clamp proteins then load the DNA polymerase onto the DNA at the primer–template junction.

The ring formed by the sliding clamp maintains the association of the polymerase with its template as replication progresses, allowing the uninterrupted synthesis of long DNA molecules. Helicases unwind the template DNA ahead of the replication fork. Single-stranded DNA-binding proteins (eukaryotic replication factor A, or RFA) then stabilize the unwound template DNA so that the single-stranded DNA can be replicated. The enzymes involved in DNA replication, in combination with their accessory proteins, synthesize both leading and lagging strands of DNA simultaneously at the replication fork.

The idea that DNA polymerases track like locomotives along the DNA template during DNA replication is pervasive and is probably based on the misperception that the polymerase is smaller than the DNA (Cook 1999). We now know that the DNA polymerase/protein complexes involved in DNA replication can be much larger than the DNA template.

An alternative model to the "movement" of polymerase along the DNA template has been proposed in which the fixed polymerase complexes "reel in their DNA templates" as they extrude newly made DNA in replication "foci" or replication factories within the cell. This "fixed" model assumes that the DNA polymerase complex is fixed and the DNA rotates around it. This is a simple solution to the potential problem of untangling DNA strands that twine around each other if the DNA polymerase moves (Cook 1999).

DNA polymerases in eukaryotes have 3′ to 5′ exonuclease activity in addition to their polymerase activity, which means that DNA polymerase can excise a misincorporated nucleotide by proofreading during DNA replication. DNA mismatch correction further minimizes replication errors by a survey of newly synthesized DNA strands. Furthermore, accessory factors such as DNA helicases apparently improve accuracy during DNA elongation, possibly because of resolution of stalled replication forks. Despite all these precautions, occasional misincorporated nucleotides or deletions and insertions may remain.

1.19. Telomeres at the End: A Solution to the Loss of DNA during Replication

Because DNA synthesis occurs exclusively in the 5′ to 3′ direction and initiation requires a short RNA primer, the extreme 5′ end of a linear DNA strand will consist of an RNA primer (Figure 1.13B). If this RNA primer is not replaced by deoxyribonucleotides, the chromosome would gradually decrease in length after each replication during mitosis, which could seriously affect gene function over time. However, linear chromosomes normally are stable because they have a specialized structure at their ends called a telomere (Zakian 1989).

Telomeres contain a series of species-specific repeated nucleotide sequences that are added to the ends of eukaryotic chromosomes by an enzyme called **telomerase**. Telomerase is a **reverse transcriptase**, meaning that it can transcribe DNA from an RNA template. A few copies of a short repetitive sequence (called the telomere sequence) are required to prime the telomerase to add additional copies to form a telomere. There are also longer, moderately repetitive nucleotide sequences subterminal to the telomere sequences.

1.20. DNA Replication Fidelity and DNA Repair

Faithful maintenance of the genome is crucial to both the individual and the species. When DNA is replicated inaccurately or is damaged by endogenous (such as water or oxygen) or exogenous factors (such as UV light, chemicals, and irradiation), death can ensue. Thus, there has been strong selection for multiple mechanisms to repair damaged DNA. Generally, the cell has two classes of mechanisms with which to repair DNA: 1) direct repair, and 2) removal of the damaged bases followed by their replacement with newly synthesized DNA (excision repair).

Direct repair Two types of damage, DNA damage caused by UV light and modifications of guanine by the addition of methyl or ethyl groups to the sixth oxygen position of the purine ring, are repaired directly (Cooper 2000).

Removal of damaged DNA components (excision repair) The most common repair mechanism in cells involves removal of damaged components of the DNA. These excision repair systems can be divided into: base-excision repair, nucleotide-excision repair, and mismatch repair. **Base-excision** repair involves removal of only the damaged base from the DNA strand. **Nucleotide-excision** repair operates mainly on damage caused by environmental mutagens and involves DNA synthesis and ligation to replace an excised oligonucleotide (Lindahl and Wood 1999). In **mismatch repair**, the mismatched bases that are incorporated during replication occasionally are not removed by the proofreading activity of DNA polymerase. The ones that are not removed are corrected by the mismatch repair system.

Postreplication repair systems Finally, if the DNA is not repaired prior to replication by the above mechanisms, a postreplication repair system comes into play. Postreplication repair (recombinational repair) can repair several types of damage to DNA, including double-strand breaks introduced into DNA by irradiation.

1.21. Mutations in the Genome

Changes in the genetic material (**genotype**) of an organism occur if DNA repair is not successful. Such changes are **mutations**. Many kinds of mutations can occur: within an exon, within introns, or in the chromosomal regions (**intergenic regions**) located between the genes. If a mutation occurs in an intergenic region, it may be silent and have no detectable effect on the cell or individual. If a mutation occurs in an exon, it may alter the protein product of the gene and cause a change in the organism's **phenotype** (or appearance). A mutation in an intron may not have an effect on the phenotype, but it can have an effect if there are regulatory elements in the intron that are important for proper gene function.

An organism with the "normal" appearance (phenotype) for that species is called the "**wild type**"; an organism with a phenotype that has been changed is a **mutant**. If the mutation is **dominant** (meaning that only a single copy is required to cause the change in phenotype), the name of the gene is capitalized. If the mutation is **recessive** (meaning that both copies of the gene carry the mutation), the name is not capitalized.

A **mutagen** is a chemical or physical agent that causes changes in bases. Mutagens include ultraviolet radiation, X-irradiation, ethyl methane sulfonate (EMS), base analogues such as 5-bromouracil, acridine dyes, and nitrous acid. Mutations occur spontaneously approximately once in every 10^8 base pairs/cell division, or they can be induced by the experimenter.

Table 1.2. Mutations Affect DNA Sequence, Gene Function, Gene Regulation, and the Phenotype of the Organism

Changes in DNA sequence	
Point mutation	Replacement of one nucleotide by another.
Transition	A point mutation in which a purine is changed to a purine (A ↔ G) or a pyrimidine to a pyrimidine (T ↔ C).
Transversion	A point mutation in which the change is purine to pyrimidine (A or G ↔ T or C).
Changes in the gene	
Silent mutation	Sequence changes in an intergenic region usually result in no phenotypic changes. Changes in a gene can be silent if a point mutation occurs in the third nucleotide of a codon which, because of the degeneracy of the code, does not alter the amino acid.
Nonsense mutation	A point mutation that alters a codon specifying an amino acid into a termination codon, which will prematurely terminate the polypeptide produced, changing the activity of the protein and altering the phenotype.
Frameshift mutation	Insertions or deletions that are not in multiples of 3 can cause changes in the amino acids downstream from the mutation, resulting in a mutant phenotype.
Changes in gene regulation	
	Mutations in regulatory genes alter the organism's ability to control expression of a gene normally subject to regulation.
Changes in the organism	
Lethal mutations	Mutations that alter the function of an essential gene product so that the organism cannot survive.
Conditional lethal	Individuals with these mutations can survive under a particular set of conditions, such as a specific temperature range, but die if reared outside these conditions.
Back mutations	Organisms sometimes revert to the wild-type phenotype after a second mutation occurs which restores the original nucleotide sequence of the mutated gene.
Reversions	Mutations can be corrected by restoring the original phenotype, but not the original DNA sequence, in the mutated gene by altering a second site within the gene.
Suppression	The effects of a mutation can be altered by a new mutation that occurs in a different gene.

Mutations affect the DNA sequence, gene organization, gene regulation, or gene function (Table 1.2). A **point mutation** is the replacement of one nucleotide by another (substitution). A substitution can be either a transition or a transversion. **Transitions** involve changes between A and G (purines) or T and C (pyrimidines), while **transversions** involve changes between a purine and a pyrimidine.

An **insertion** or **deletion** is the addition or deletion of one or more nucleotides. An **inversion** is the excision of a part of the DNA molecule followed by its reinsertion into the same position but with a reversed orientation. It has been discovered that an inversion in *Drosophila buzzatii* was caused by a transposable element called *Galileo*, and this may be the mechanism by which many inversions occur (Caceres et al. 1999).

Some mutations are lethal, while others have an effect on the organism that can range from phenotypically undetectable (silent) to lethal only under certain circumstances (**conditional lethal**). For example, a number of mutations are temperature sensitive, and the organism can survive if reared within one temperature range but will die if reared at higher temperatures.

A **silent mutation** may occur if the third base in a codon is altered but, because the genetic code is degenerate, there is no change in the amino acid specified. These also are called **synonymous** mutations. There is no change in protein structure or function from a silent mutation.

Some changes in codons alter the amino acid specified. Called **nonsynonymous mutations**, these are point mutations that result in changes in the amino acid. Most point mutations that occur at the first or second nucleotide positions of a codon will be missense, as will a few third-position changes. A polypeptide with an amino acid change may result in a changed phenotype, depending on the precise role the altered amino acid plays in the structure or function of the polypeptide. Most proteins can tolerate some changes in their amino acid sequence if the alteration does not change a segment of the polypeptide essential for the structure or function of the protein.

Nonsense mutations are point mutations that change a codon specifying an amino acid into a termination codon, which will produce a truncated gene which codes for a polypeptide that is terminated prematurely. In many cases, essential amino acids will be deleted and the protein's activity will be altered, resulting in a mutant phenotype.

Frameshift mutations result if addition or deletion of base pairs occurs that is not in a multiple of three. The polypeptide produced will likely have a complete new set of amino acids produced downstream of the frameshift. Frameshifts usually produce mutant phenotypes.

Occasionally, **back mutations** may occur to reverse a point mutation. **Reversions** sometimes occur when the original phenotype is restored by a new change in the nucleotide sequence. In reversions, the original mutation is not restored to its previous unmutated form; rather, the second mutation restores the code for the original amino acid because the code is degenerate. **Regulatory mutations** are mutations that affect the ability to control expression of a gene.

The movement of a **transposable element** into a gene can also create mutations in genes. Transposable elements (TEs) are segments of foreign DNA that can move into genomes. When TEs move into a gene, as is shown in Figure 1.15, the gene will be inactivated or the gene product will be altered and produce a visible phenotype (mutation). Transposable elements can cause other types of mutations, including inversions (Caceres et al. 1999). Transposable elements are found in most eukaryotic organisms, and there are many types. Transposable elements are important for understanding genome evolution and for genetic engineering and are discussed further in Chapters 2, 8, and 14.

1.22. Common Conventions in Genetic Terminology

A **wild-type gene** is normally identified only after a mutation has disrupted the phenotype of an organism. Mutations commonly are given a descriptive name, such as "white eyes." The name of the gene usually is italicized (*white*) and is abbreviated using one, two, or three italicized letters (*w*). If the mutation is dominant, the name and abbreviation are capitalized (*White*); they are in lowercase (*white*) if the

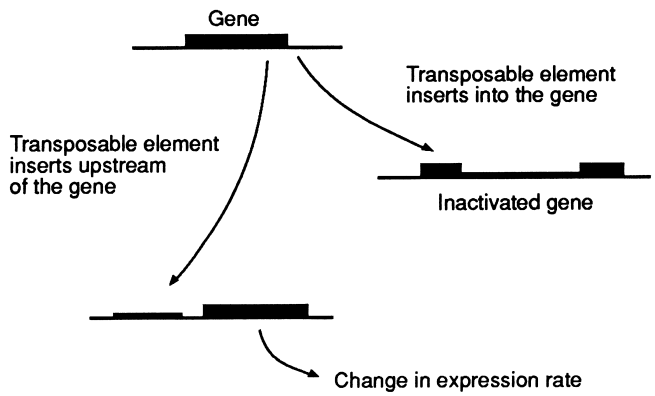


Figure 1.15. Movement (transposition) of transposable elements into chromosomes can result in mutations that inactivate genes or alter their expression.

mutation is recessive. Individuals that are homozygous for the recessive *w* mutation are *w/w* and have white eyes. Heterozygous flies are w/w^+, with the wild-type allele designated as w^+, and their appearance (phenotype) should be wild type. The gene product is called the white product or white protein and is not italicized. The term for the gene product may be abbreviated as the w protein. Sometimes the protein product is designated by the gene name but is capitalized to distinguish it from the gene (WHITE).

1.23. Independent Assortment and Recombination during Sexual Reproduction

For organisms to survive and evolve with changing environmental conditions, they need to be able to generate genetic variability. Mutations are one source of genetic variability and thus are not always undesirable. Another source of genetic variability is the result of sexual reproduction.

In sexually reproducing organisms, the progeny produced by parents that have different versions of genes (different alleles, AA or aa) will have a different combination of alleles. This shuffling of the genetic information during sexual reproduction is due to the **independent assortment** of homologous chromosomes into the gametes during **meiosis**. Thus, an individual of genotype AaBb, in which the genes A and B are located on different chromosomes, will produce equal numbers of four different types of gametes: AB, Ab, aB, or ab.

Crossing over also leads to recombination between DNA molecules. Crossing over occurs between homologous chromosomes during the production of eggs or sperm in meiosis I and results in an exchange of genetic material. Crossing over allows new combinations of different genes that are linked (located on the same chromosome). Thus, if a parent has two homologous chromosomes, one with A and B on the same chromosome, and the other with a and b on the homologous chromosome, a physical exchange between the chromatids during meiosis I can lead to gametes that have the following combinations: A and B; A and b; a and B; and a and b. Nonhomologous recombination, crossing over between DNA lacking sequence homology, may also occur, but the mechanism(s) involved are not understood. Meiosis and mitosis are described further in Chapter 3.

General References

Beebee, T., and J. Burke (1988). *Gene Structure and Transcription*. IRL Press, London.

Brown, T. A. (1989). *Genetics, A Molecular Approach*. Chapman and Hall, London.

Cold Spring Harbor Symposia on Quantitative Biology (2001). *Biological Responses to DNA Damage*. Vol. LXV. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cooper, G. M. (2000). *The Cell. A Molecular Approach*, 2nd Ed. ASM Press, Washington, DC.

DePamphilis, M., Ed. (1999). *Concepts in Eukaryotic DNA Replication*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Gesteland, R. F., Ed. (1999). *The RNA World*, Second Edition, Monograph 37. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hartl, D. L. (1991). *Basic Genetics*, 2nd Ed. Jones and Bartlett Publishers, Boston.

Watson, J. D., N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner (1987). *Molecular Biology of the Gene*, Vols. I and II, 4th Ed. Benjamin/Cummings, Menlo Park, CA.

References Cited

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, and S. E. Scherer et al. (2000). The genome sequence of *Drosophila melanogaster. Science* **287**: 2185–2195.

Avery, O. T., C. M. Macleod, and M. McCarty (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* **79**: 137–158.

Baker, T. A., and S. P. Bell (1998). Polymerases and the replisome: machines within machines. *Cell* **92**: 295–305.

Berezney, R., D. D. Dubey, and J. A. Huberman (2000). Heterogeneity of eukaryotic replicons, replicon clusters and replication foci. *Chromosoma* **108**: 471–484.

Bogan, J. A., D. A. Natale, and M. L. DePamphilis (2000). Initiation of eukaryotic DNA replication: conservative or liberal? *J. Cell. Physiol.* **184**: 139–150.

Caceres, M., J. M. Ranz, A. Barbadilla, M. Long, and A. Ruiz (1999). Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**: 415–418.

Cook, P. R. (1999). The organization of replication and transcription. *Science* **284**: 1790–1795.

Cooper, G. M. (2000). *The Cell: A Molecular Approach*, 2nd Ed. ASM Press, Washington, DC.

DePamphilis, M. L. (1999). Replication origins in the metazoan chromosomes: fact or fiction? *BioEssays* **21**: 5–16.

DeSouza, S. J., M. Long, R. J. Klein, S. Roy, S. Lin, and W. Gilbert (1998). Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. USA* **95**: 5094–5099.

DiGiulio, M. (1997). On the RNA world: evidence in favor of an early ribonucleopeptide world. *J. Mol. Evol.* **45**: 571–578.

Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**: 919–929.

Gavin, K. A., M. Hidaka, and B. Stillman (1995). Conserved initiator proteins in eukaryotes. *Science* **270**: 1667–1671.

Gilbert, W., S. J. DeSouza, and M. Long (1997). Origin of genes. *Proc. Natl. Acad. Sci. USA* **94**: 7698–7703.

Herbert, A. (1996). RNA editing, introns and evolution. *Trends Genet.* **12**(1): 6–8.

Huberman, J. A. (1995). Prokaryotic and eukaryotic replicons. *Cell* **82**: 535–542.

Jeffares, D. C., A. M. Poole, and D. Penny (1998). Relics from the RNA world. *J. Mol. Evol.* **46**: 18–36.

Johnston, W. K., P. J. Unrau, M. S. Lawrence, M. E. Glasner, and D. P. Bartel (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**: 1319–1325.

Knight, R. D., L. F. Landweber, and M. Yarus (2001). How mitochondria redefine the code. *J. Mol. Evol.* **53**: 299–313.

Lang, F., M. W. Gray, and G. Burger (1999). Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **22**: 351–397.

Leipe, D. D., L. Aravind, and E. V. Koonin (1999). Did DNA replication evolve twice independently? *Nucleic Acids Res.* **27**: 3389–3401.

Lindahl, T., and R. D. Wood (1999). Quality control by DNA repair. *Science* **286**: 1897–1905.

Logsdon, J. M., Jr., M. G. Tyshenko, C. Dixon, J. D. Jafari, V. K. Walker, and J. D. Palmer (1995). Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc. Natl. Acad. Sci. USA* **92**: 8507–8511.

Maienschein, J. (1992). Gene: historical perspectives. pp. 122–127. In: *Keywords in Evolutionary Biology*, E. Fox-Keller, and E. A. Lloyd, Eds. Harvard Univ. Press, London.

Muller, H. J. (1947). The gene. *Proc. R. Soc. London B* **184**: 1–37.

Nelkin, D. (2001). Molecular metaphors: the gene in popular discourse. *Nat. Rev. Genet.* **2**: 555–559.

Nesbo, C. L., Y. Boucher, and W. F. Doolittle (2001). Defining the core of nontransferable prokaryotic genes: the Euryarchaeal core. *J. Mol. Evol.* **53**: 340–350.

Poole, A. M., D. C. Jeffares, and D. Penny (1998). The path from the RNA world. *J. Mol. Evol.* **46**: 1–17.

Prusiner, S. B., and M. R. Scott (1997). Genetics of prions. *Annu. Rev. Genet.* **31**: 139–175.

Sutton, M. D., and G. C. Walker (2001). Managing DNA polymerases: coordinating DNA replication, DNA repair, and DNA recombination. *Proc. Natl. Acad. Sci. USA* **98**: 8342–8349.

Trotman, C. N. A. (1998). Introns-early: slipping lately? *Trends Genet.* **14**: 132–134.

Tuite, M. F. (2000). Sowing the protein seeds of prion propagation. *Science* **289**: 556–557.

Tyshenko, M. G., and V. K. Walker (1997). Towards a reconciliation of the introns early or late views: triosephosphate isomerase genes from insects. *Biochim. Biophys. Acta* 1**353**: 131–136.

Watson, J. D., and F. H. C. Crick (1953). Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**: 737–738.

Zakian, O. A. (1989). Structure and function of the telomeres. *Annu. Rev. Genet.* **23**: 579–604.

2

Transcription, Translation, and Regulation of Eukaryotic DNA

2.1. Overview

Genetic information in an organism is expressed in three classes of genes: 1) structural (protein-coding) genes that are transcribed into mRNA and translated into polypeptides; 2) genes that code for ribosomal RNA, transfer RNA, or other small RNAs, in which the transcription product is used directly; and 3) regulatory sites that are not transcribed but serve as recognition sequences for proteins involved in DNA replication, transcription, and repair.

Protein-coding genes in eukaryotic organisms are transcribed from DNA into pre-messenger RNAs which then are processed into messenger RNA (mRNA). Processing of mRNA involves splicing to remove introns; mRNA also is capped and methylated at the 5′ end, and most mRNAs are polyadenylated at the 3′ end. The information in the mRNA then is translated into proteins via protein synthesis in ribosomes.

A ribosome begins protein synthesis once the 5′ end of a mRNA is inserted into it. A lengthening polypeptide chain is produced, and, once the 5′ end of the mRNA emerges from a ribosome, the mRNA can attach to a second ribosome and a second identical polypeptide can be synthesized. The assembly of amino acids into a peptide starts at the amino end and finishes at the carboxyl end. Amino acids are carried to the ribosome by transfer RNAs (tRNAs). The tRNAs are held so their anticodons form base pairs with complementary codons of the mRNA. One tRNA may recognize more than one codon because the genetic code is degenerate. According to the "wobble hypothesis," the first two bases of the mRNA codon pair according to base-pairing rules, but the third base may pair with any one of several bases.

Protein synthesis occurs on the ribosomes, located in the endoplasmic reticulum of the cell cytoplasm. Once proteins are produced they are transported into the Golgi apparatus, where they are processed and transported to their ultimate destination. The protein must be folded and, sometimes, assembled into multiprotein complexes. Folding may require the assistance of other proteins called molecular chaperones.

Gene regulation in insects and other eukaryotes is complex, diverse, and the subject of intensive research. Genes may be amplified or rearranged to yield increased gene products. Genes may be regulated by being methylated. Transcription is influenced by activator proteins, hormones, and enhancers. Alternative splicing, alternative promoters, and translational control are employed in gene regulation. Insulators or boundary elements are naturally occurring DNA sequences that protect genes from position effects, establishing independent functional domains within the chromosome.

2.2. Introduction

The Central Dogma, that DNA is transcribed into RNA which subsequently is translated into proteins, describes the process by which information contained in the DNA is made available to the cell and organism (Figure 1.1).

The proteins specified by the genetic code have many functions in the cell. Structural proteins form part of the framework of the organism, such as the sclerotin in the exoskeleton of insects. Contractile proteins enable organisms to move. Catalytic proteins, or enzymes, regulate the diverse biochemical reactions taking place within the cell. Transport proteins carry important molecules throughout the body. Regulatory proteins control and coordinate biochemical reactions in the cell and the organism as a whole. Protective proteins

(antibodies) protect against infectious agents and injury. Storage proteins store products for future use.

The development of a functioning organism involves the coordinated activity of a large number of different proteins, the information for which is encoded in the genes (DNA). In addition, genes carry the code for ribosomal RNA (rRNA), transfer RNA (tRNA), and other small RNAs. Ribosomal, transfer RNA, and small RNA molecules are used directly without being translated into proteins.

Research to decipher how the genetic information in the DNA is utilized by the cell and organism is complex and rapidly advancing (Herbert and Rich 1999, Lee and Young 2000). Much of the early work on transcription and translation of genetic information was conducted using prokaryotes.

Eukaryotes, which include insects, differ from prokaryotes in several important ways. First, eukaryotic genes generally are located in more than one chromosome. Eukaryotes are genetically more complex (with perhaps 15,000 to 40,000 genes). Furthermore, most eukaryotic genes that code for proteins are split, with one or more noncoding **introns** interspersed among the coding **exons**.

Control elements, such as **promoters** and **enhancers**, are important components of gene regulation in eukaryotes. Furthermore, within a particular tissue, a mixture of active and inactive genes is present on each chromosome. Recent research in *Drosophila* indicates that specific sequences called **boundary elements** insulate the active from the inactive regions of the chromosomes (Bell and Felsenfeld 1999, Bell et al. 2001).

Finally, the nuclear membrane in eukaryotes separates the processes of transcription in the nucleus and translation in the cytoplasm in both time and space. The intricacies of transcription and translation of eukaryotic DNA are still being unraveled.

2.3. RNA Synthesis Is Gene Transcription

Transcription is the first stage of gene expression (Figure 2.1). During transcription, the coding strand of DNA serves as a template for synthesis of an RNA molecule. The sequence of the RNA molecule is determined by complementary base pairing so that the RNA is a complementary transcript (copy) of the coding strand of DNA.

Transcription requires four ribonucleoside 5′-triphosphates: ATP, GTP, CTP, and UTP (recall that uracil substitutes for thymine in RNA). A sugar–phosphate bond is formed between the 3′-OH group of one nucleotide and the 5′-triphosphate of a second nucleotide by the enzyme RNA polymerase. Unlike DNA polymerase, RNA polymerase can initiate RNA synthesis without requiring a primer. The sequence of bases in the RNA molecule is determined by the sequence of bases in the DNA coding strand. Base pairing occurs between the DNA bases and the newly forming single-stranded RNA molecule. Nucleotides are added to the 3′-OH end of the growing end of the RNA molecule, and thus synthesis of RNA proceeds in the 5′ to 3′ direction, as does DNA synthesis (Figure 2.2).

Eukaryotes have three types of nuclear RNA polymerase, called RNA polymerase I, II, or III, and each is responsible for transcribing the three different classes of genes. RNA polymerase I primarily is responsible for synthesis of Class I genes, which includes the large ribosomal RNAs (rRNA) in the nucleolus. Class II genes include all the DNA sequences that code for proteins and some small nuclear RNAs (snRNAs). Class III genes include the transfer RNA (tRNA) genes, 5S ribosomal RNA (rRNA) genes, and genes for some small nuclear RNAs.

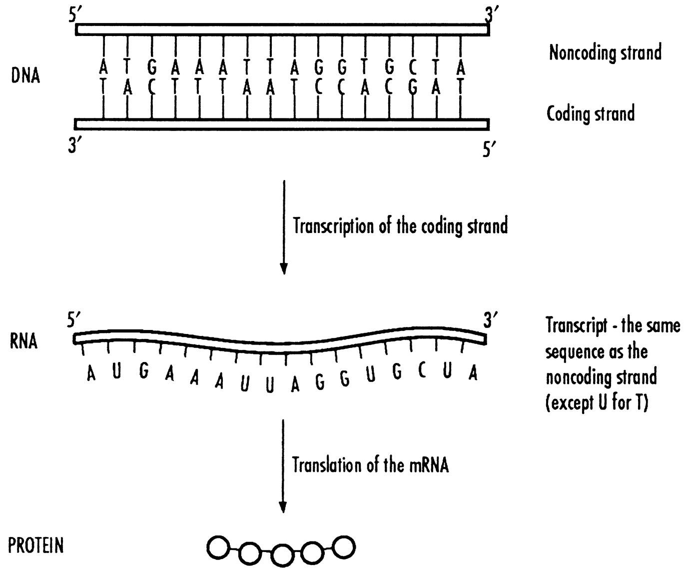


Figure 2.1. Protein-coding gene expression involves transcription of the coding strand of DNA to pre-messenger RNA, which is then processed to messenger RNA, which is then translated into proteins.

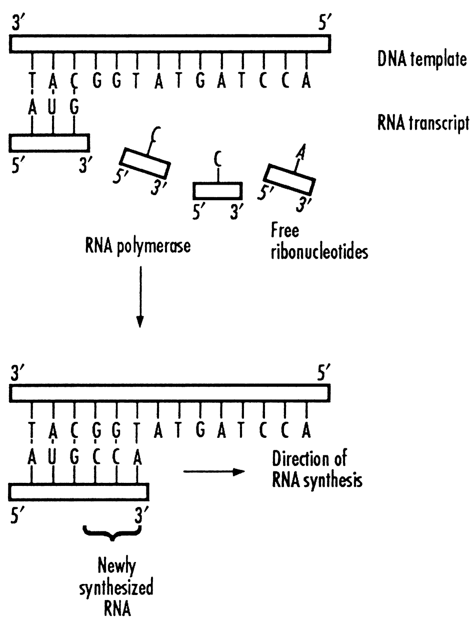


Figure 2.2. RNA synthesis involves polymerization of free ribonucleotides by an RNA polymerase in the 5′ to 3′ direction. Thus, the DNA template is read in the 3′ to 5′ direction.

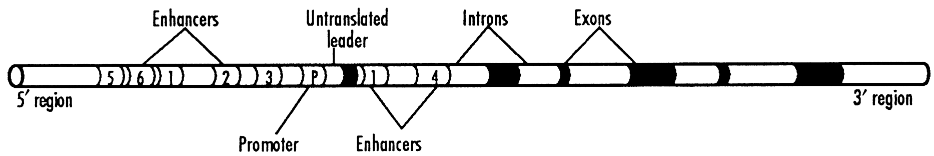


Figure 2.3. Components of a typical Class II eukaryotic gene that codes for proteins include: noncoding introns that are spliced out of the pre-messenger RNA, coding exons (shown in black), a promoter to which RNA polymerase II attaches to initiate RNA synthesis, and several enhancers (here numbered 1–6) that influence gene regulation.

Figure 2.3 illustrates several of the elements of a typical eukaryotic Class II gene. This gene consists of noncoding introns, amino acid-coding exons, one or more promoters, and several (in this example, six) enhancer elements.

2.4. Transcription Involves Binding, Initiation, Elongation, and Termination

RNA polymerase II synthesizes all messenger RNA in eukaryotes and is a large and complex molecule (Cramer et al. 2000, Lee and Young 2000). RNA polymerase II can unwind DNA, polymerize RNA and proofread the developing transcript. RNA polymerase II recognizes promoters and responds to regulatory signals (Cramer et al. 2000).

Transcription is a complex, multistep process involving binding of RNA to the DNA, initiation, elongation, and production of an mRNA transcript (Hoffman et al. 1997, John and Workman 1998). To initiate transcription of a Class II gene, binding of RNA polymerase II must occur at a specific point upstream of the DNA to be transcribed. This usually involves loading a protein called TFIID onto a promoter, followed by the recruitment of a group of proteins called general transcription factors (GTFs). These promoter-bound GTFs then recruit RNA polymerase II to form a preinitiation complex.

The specific attachment sites of the preinitiation complex are called **promoters** and are typically 20 to 200 nucleotides long. Different eukaryotic promoter sequences are known, but certain common, or consensus, patterns occur. For example, many protein-coding genes contain the promoter sequences TATAAT and CAAT (Figure 2.4), often called the TATA and CAAT boxes. The location of the TATA sequence may vary, and not all genes have the TATA sequence. Housekeeping genes (genes that are expressed in all cells in order to maintain fundamental activities) may lack the TATA box and have a GC-rich region about 33 nucleotides upstream from the start site. The actual sequences of the promoter vary from gene to gene; the "strength" of the promoter affects the extent to which each gene is expressed.

Eukaryotic promoters differ from prokaryotic ones by having other types of DNA sequences called **enhancers** that influence the efficiency with which RNA polymerase II, and accessory factors can assemble at a promoter to initiate transcription of the DNA (Figures 2.3, 2.4). Enhancers can be at a great distance relative to the RNA start site and can be upstream or downstream of the gene (Blackwood and Kadonaga 1998). For example, the enhancer of the *Drosophila cut* gene is 85 kb upstream from the promoter. Enhancer sequences vary in size from 50 base pairs (bp) to 1.5 kilobases (kb). Enhancers activate their target gene in a specific cell type at a particular stage in development. Once RNA

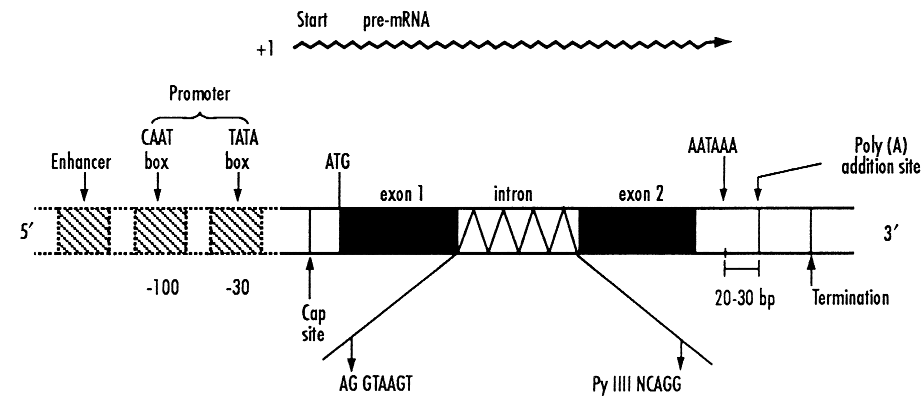


Figure 2.4. A more detailed view of a eukaryotic class II gene that codes for proteins. Promoters often have CAAT and TATA boxes upstream from the start site. The left junction (splice donor, AG GTAAGT) and right junction (splice acceptor, NCAGG) sequences of the intron are shown. Splice sites are indicated by the arrow between AG and GT of the splice donor and between the two Gs of the splice acceptor. Pre-messenger RNA synthesis is initiated at the +1 (start) site and proceeds in the 3′ direction.

polymerase recognizes the specific attachment site, the next two phases in transcription can occur: initiation and elongation (John and Workman 1998).

Initiation involves "melting" the DNA around the start site of transcription. The transition from an initial complex to a transcribing complex is accompanied by structural changes and movement of the DNA (Cramer et al. 2000).

2.5. RNA Transcripts Are Longer Than the Protein-Coding Gene

The actual RNA transcript produced in eukaryotes is longer than the gene it is transcribing because RNA polymerase transcribes a **leader sequence**, the length of which varies from gene to gene. When the end of the gene has been reached, RNA polymerase continues to transcribe a **trailer segment** before terminating its activities.

Termination of class II (protein-coding) genes appears to occur hundreds or even thousands of nucleotides (nt) downstream of the 3′-end of the mRNA, which in turn generally lies about 35 nt downstream from the site coding for a polyadenylation (polyA) signal, AAUAAA (Figure 2.4).

The number of noncoding introns in class II genes varies, as does their length. The boundaries between introns and exons often are determined by a consensus sequence to ensure that the introns are spliced out of the transcript in a precise manner (Figure 2.5).

2.6. RNA of Protein-Coding Genes Must Be Modified and Processed in Eukaryotes

In prokaryotes, the RNA transcript can be translated immediately into specific amino acid sequences. In other words, it is the messenger RNA (mRNA). In eukaryotes, however, RNA

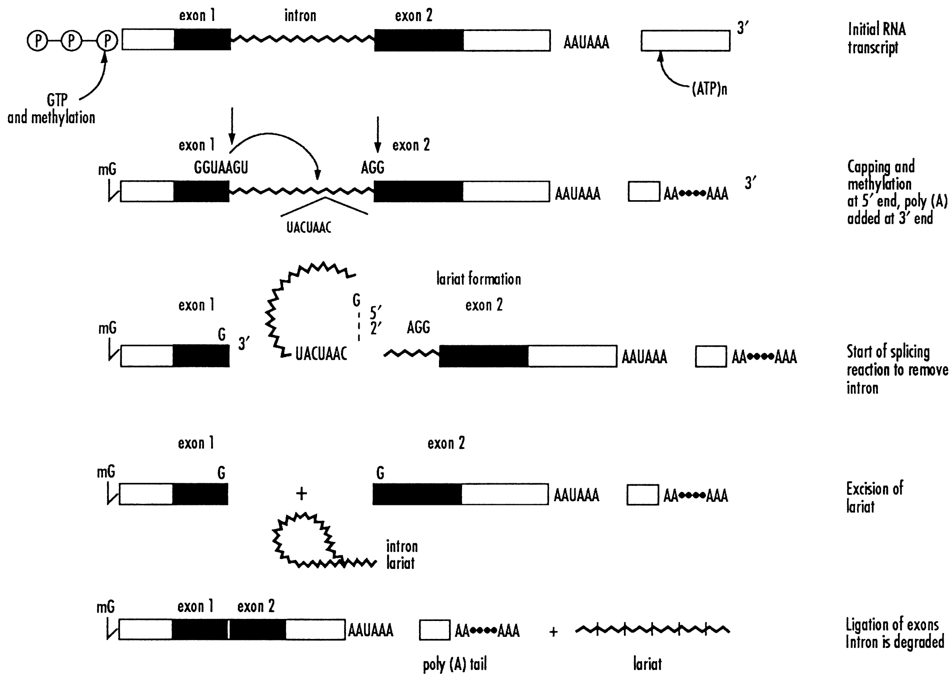


Figure 2.5. Pre-mRNA is processed before it becomes mRNA. The initial transcript is capped at the 5′ end by adding a G, and the G is then methylated. The 3′ end is polyadenylated after the signal AATAAA. The intron is removed after the left end is cut, and a lariat is formed between the G at the 5′ end of the intron and the 3′ A nucleotide with an unusual 5′–2′ phosphodiester bond. The right end of the intron is cut, the lariat is released, and the exons are then joined. The released lariat is degraded later.

transcribed from DNA must be modified and processed before it can function as mRNA (Figure 2.5).

Processing the pre-messenger RNA involves two activities: 1) modifying both ends of the RNA molecule, and 2) excising the noncoding sequences (introns) contained within the coding region.

Modifying the ends involves **capping** the 5′ end and adding a polyA tail to the 3′ end. Newly synthesized eukaryotic RNA molecules are capped at the 5′ end by adding a terminal guanine (G) that has been methylated on the 7 position and linked to the start site by an unusual 5′–5′ triphosphate linkage (Figure 2.5). Capping appears to be necessary to enable the ribosome to bind with the mRNA before protein synthesis can begin. The methylated G nucleotide is added in a two-step process, with methylation occurring after a standard G has been added. In some eukaryotes, additional methyl groups may be added to one or both of the next two nucleotides of the mRNA molecule.

The 3′ end of eukaryotic RNA is modified by adding 40 to 200 adenine (A) residues to a region near the 3′ end of the transcript to produce the **poly(A) tail** (Figure 2.5). The polyadenylation does not simply add the A residues to the end of the transcript. First, a cleavage occurs between 10 and 30 nt downstream of a specific polyadenylation signal, which in insects is usually AAUAAA and is found in the 3′ noncoding region of the RNA. This results in an intermediate 3′ end to which the poly(A) tail is added by the enzyme

poly(A) polymerase. The length of the poly(A) tail may determine how long the mRNA survives in the cytoplasm before being degraded.

2.7. Splicing Out the Introns

The third modification of the pre-mRNA involves splicing to remove any introns. Splicing takes place in two steps. Introns have a 5′ donor and a 3′ acceptor end with common consensus sequences (Figure 2.5). The 5′ donor end typically has the sequence GGUAAGU. After a cut in the donor site, the G at the 5′ end forms a loop by attaching to an A nucleotide a short distance upstream from the pyrimidines near the acceptor splice site. The consensus sequence of the 3′ acceptor site is AGG. In the final step, a cut is made in the acceptor site and the intron is freed. The exons are then joined together. The excised loop (intron) is released as a lariat-shaped structure and is later degraded (Figure 2.5).

RNA splicing occurs in large multicomponent complexes called **spliceosomes**. Spliceosomes are composed of more than 50 proteins and five types of small nuclear RNA (snRNA) molecules (U1, U2, U4, U5, and U6) (Reed 2000). The active catalytic components of the spliceosome are the snRNAs rather than the proteins, although the proteins are required and participate in both assembly of the spliceosome and the splicing reaction. In addition, a number of proteins play auxiliary roles in splicing and spliceosome assembly.

Because many pre-mRNAs contain multiple introns, the splicing machinery must be able to identify and join the appropriate 5′ end and 3′ splice sites to produce a functional mRNA. The specificity of the splicing operation is determined by the snRNAs, which contain sequences that are complementary to the splice junctions (Cooper 2000, Hastings and Krainer 2001).

Introns often have no function, and synthetic genes lacking introns can function quite well. However, some introns are important in gene regulation and determine when, or in what tissue, the gene will be transcribed. For example, sex determination in *D. melanogaster* depends on a cascade of splicing and the pre-mRNA of the *double-sex*$^+$ gene of the female contains exon 4, while males lack it. Likewise, the splicing of *Sex lethal*$^+$ and *transformer*$^+$ varies by sex (see Chapter 10).

Mutations in introns can be neutral or alter gene regulation. Mutations in the splicing signals may result in two classes of mutations. If an intron is not spliced out, a mutant protein can be produced that functions abnormally. If splicing occurs at a different site than normal, an abnormal mRNA is produced and a mutant protein also will be produced.

More than half the introns in *Drosophila* and other invertebrates are less than 80 nt long (Guo and Mount 1995). Once mRNA is produced, it must be transported through the nuclear envelope to the cytoplasm where it is translated.

2.8. Translation Involves Protein Synthesis

Translation is the second stage of Class II gene expression in which the information in the mRNA is used to direct the synthesis of a polypeptide, the amino acid sequence of which is determined by the nucleotide sequence of the RNA.

The genetic code consists of a triplet of adjacent ribonucleotides that specify an amino acid (Table 1.1). Translation requires ribosomes, transfer RNA, a set of enzymes to catalyze the attachment of each amino acid to its corresponding tRNA molecule (aminoacyl

tRNA synthetases), and initiation, elongation, and termination factors. Translation occurs in ribosomes located in the cytoplasm.

Ribosomes are cellular organelles, consisting of two subunits, each composed of ribosomal RNA and proteins. The smaller subunit binds mRNA and the anticodon end of tRNAs and helps to decode the mRNA. The larger subunit interacts with the amino acid-carrying end of tRNAs and catalyzes the formation of the peptide bonds.

The mechanisms whereby ribosomes engage a mRNA and select the start site for translation are more complicated in eukaryotes than in prokaryotes (Dever 1999, Kozak 1999, Preiss and Hentze 1999). However, the fundamental components of translation are conserved: ribosomal RNAs are strongly conserved in both primary and secondary structure among all organisms. The majority of ribosomal proteins are conserved, as well as the elongation factors, the tRNAs, and the aminoacyl-tRNA synthetases (Kyrpides and Woese 1998).

Initiation sites in eukaryotic mRNAs are reached by a scanning mechanism that predicts translation should start at the AUG codon nearest the 5′ end of the mRNA. The selection of the start codon sets the reading frame that is maintained throughout subsequent steps in the translation process. Protein synthesis is often regulated at the level of initiation, which makes it an important step.

Recent evidence indicates that the RNA of the larger ribosomal subunit carries out the formation of the peptide bonds; thus, the ribosome is a **ribozyme**, meaning that the RNA acts as an enzyme. The RNA carries out the key peptidyl transferase reaction (Cech 2000, Nissen et al. 2000). By contrast, the proteins in the ribosome are structural units and help to organize key catalytic RNA elements.

Translation of the genetic information in eukaryotes begins when a mRNA molecule binds to the surface of a ribosome and the initiation codon (AUG) is selected. The Met-tRNA interacts with the AUG start codon of the mRNA. More than nine eukaryotic initiation factors (elFs) have been identified, and several of them are composed of multiple polypeptide chains. This large number of polypeptides suggests that protein-protein interactions play an important role in initiation of translation (Dever 1999).

Current models of initiation of translation suggest that factor elF2 forms a complex with GTP and the initiator Met-tRNA (Dever 1999). This complex then associates with a 40S ribosomal subunit and several other initiation factors (including elF1, elF1A, and elF3). This preinitiation complex is then ready to bind to an mRNA that has been prepared for ribosome binding by the elF4 proteins. The ribosome is thought to bind to the mRNA near its 5′ end and then scan the mRNA for the initiating AUG codon in order to form the 48S preinitiation complex. The elF5 factor stimulates GTP hydrolysis by elF2 in the 48S complex, which results in release of the initiation factors from the ribosome and enables the 60S ribosomal subunit to complete the initiation process. The ribosome is now poised for translation elongation. Hydrolysis of GTP in the 48S complex results in the release of a stable elF2–GDP complex.

Transfer RNAs carry an amino acid to the ribosome where they bind to the mRNA molecule attached to the ribosome. Transfer RNAs have a three-base sequence, called an **anticodon**, that is complementary to a specific codon in mRNA and a site to which a specific amino acid is bound (Figure 2.6). Binding between the mRNA and tRNA occurs by hydrogen bonds. Proteins within the ribosome function as cofactors, buttressing, stabilizing, and orienting the floppy ribosomal RNA into a specific, active ribozyme.

Peptide bonds are made between the successively aligned amino acids until the stop codon at the end of the mRNA is reached (UAA, UAG, or UGA) and the completed

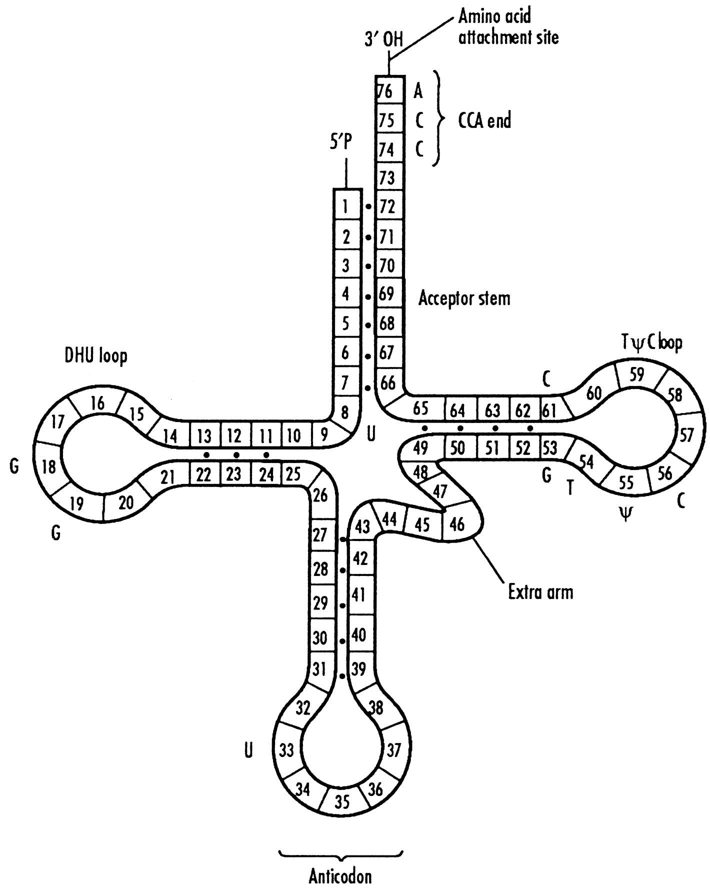


Figure 2.6. A tRNA molecule has a complex shape. Bases numbered from 1 to 76. A few bases that are present in almost all tRNA molecules are identified by letters. The Greek letter ψ is a symbol for the unusual base pseudouridine. The amino acid attachment site is at the 3′ end of the tRNA, and the anticodon (the genetic code that determines the amino acid order) is at the bottom of the diagram.

protein is released. The polypeptide is thus synthesized from the amino end toward the carboxyl end.

Transfer RNAs are small, single-stranded molecules ranging in size from 70 to 90 nucleotides (Figure 2.6). Internal complementary base sequences allow the molecule to form short double-stranded regions, which yields a folded molecule in which open loops are connected to each other by double-stranded stems. The three-dimensional structure of tRNA molecules is complex. One significant region is the anticodon sequence region, which consists of three bases that can base pair with the codon in the mRNA. A second critical site is the 3′ end of the molecule where the amino acid attaches.

A specific enzyme called **aminoacyl tRNA synthetase** matches each amino acid with the tRNA attachment site. Transfer RNA molecules and their synthetases are designated by giving the name of the amino acid that is specific to each particular tRNA molecule. Thus, leucyl-tRNA synthetase attaches leucine to $tRNA^{Leu}$. If an amino acid

is attached to a tRNA molecule, it is "charged." Usually, one, and only one, aminoacyl synthetase is found for each amino acid. However, there are fewer aminoacyl synthetases than there are codons for amino acids. Thus, the aminoacyl synthetases must recognize more than one codon. The **wobble hypothesis** suggests that base pairing is most critical with the first two bases of the codon but that pairing is extremely flexible in the third position.

Gene translation in eukaryotes usually involves structures more complex than a single ribosome processing a single mRNA molecule. Thus, after about 25 amino acids have been joined together in a polypeptide, the AUG initiation codon is free of the ribosome and a second polypeptide can begin to form. When the second ribosome has joined about 25 amino acids, a third ribosome can attach to the initiation site. This can result in mRNA molecules with many ribosomes, all moving in the same 5′ to 3′ direction. This large unit is called a **polysome**. Figure 2.7 illustrates a polysome isolated from a midge larva, *Chironomus*. This electron micrograph, magnified ca. 140,000 times, shows the start of a mRNA molecule on the bottom right (Kiseleva 1989). The structure at the top shows the end of the molecule, with the growing proteins shown attached to the ribosomes.

Proteins are linear chains of amino acids that adopt unique three-dimensional structures which allow them to carry out their biological functions. All the information needed to specify a protein's three-dimensional structure is contained within its amino acid sequence (Denton and Marshall 2001). Protein folds are the basic constructional units of proteins, each consisting of between 80 and 200 amino acids. Some proteins consist of a single fold, but most are a combination of two or more. Protein folds are limited to a finite number of distinct structural families containing a number of closely related forms (Denton and Marshall 2001). It is estimated that the total number of permissible folds is about 4000. Given suitable conditions, most small proteins will fold spontaneously into their three-dimensional form (Baker 2000). The ability to predict protein structure and folding mechanisms should help in understanding how the protein functions. The fundamental physics underlying folding appears to be relatively simple and is determined by the topology of the folded protein (Baker 2000). The challenge is to improve protein-folding models so that they can contribute to the understanding of genome sequence information.

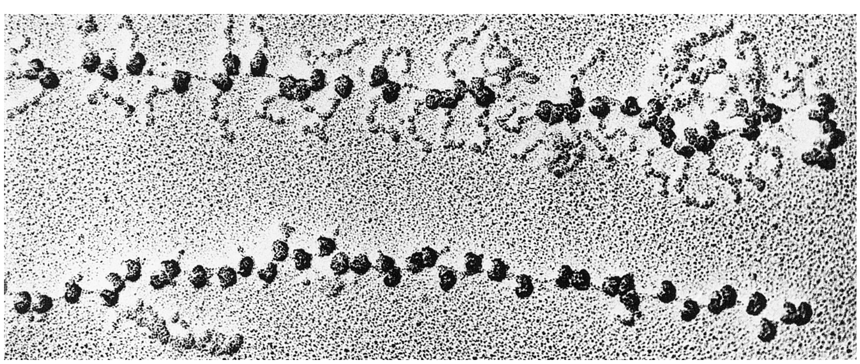


Figure 2.7. A micrograph of polysomes from salivary gland cells of the midge *Chironomus*. Here the ribosomes are moving in order along messenger RNA, gradually extending their individual protein chains. The start of the polysome is at the bottom and the end is at the top.

2.9. mRNA Surveillance: Damage Control

It is very important that mRNA be accurately produced; otherwise, damaged or truncated proteins are produced (Hilleren and Parker 1999, Maquat and Carmichael 2001). An **mRNA surveillance** system in eukaryotes monitors pre-mRNA processing and RNA translation. mRNA surveillance increases the fidelity of gene expression by degrading aberrant mRNAs that, if translated, would produce truncated proteins.

RNA surveillance also assesses whether the transcripts possess the proper regulatory elements. Thus, mRNA surveillance is important in defining which mRNA molecules become available for translation and, hence, which proteins are produced within the cell.

The actual events in mRNA surveillance involve a rapid decapping of the transcript leading to 5′ to 3′ exonucleolytic degradation. Transcripts targeted for decay by the mRNA surveillance system are decapped whether or not the 3′ poly(A) tail is shortened.

Messenger RNA is recognized as imperfect only after the mRNA has been translated by the ribosomes. Apparently, a critical step in the surveillance process is the recognition that the mRNA has an improper termination signal. Translation termination at a nonsense or premature termination codon is somehow interpreted as improper (Hilleren and Parker 1999).

2.10. Import and Export from the Nucleus

The nucleus distinguishes eukaryotes from prokaryotes. The nucleus contains the nuclear genetic information and is the cell's control center where DNA replication, transcription, and RNA processing take place. The final stage of gene expression, translation, occurs in the cytoplasm.

The nucleus is surrounded by a nuclear envelope consisting of two nuclear membranes constructed of phospholipid bilayers, an underlying lamina (a network of lamin filaments that provide structural support), and nuclear pore complexes. The nuclear lamina is thought to serve as a site where the chromosomes attach. The inner and outer membranes of the nuclear membrane are joined at the nuclear pore complexes, which are the only channels through which small polar molecules and macromolecules are able to travel through the nuclear envelope.

The nuclear membranes prevent the free movement of molecules between the nucleus and the cytoplasm, thus maintaining the nucleus as a distinct biochemical compartment. The outer nuclear membrane is similar in structure to the membranes of the endoplasmic reticulum, to which it attaches.

The **nuclear pore** complex, through which all molecules entering and exiting the nucleus must pass, has a diameter of about 120 nanometers (nm) and a mass of approximately 125 million daltons, which is about 30 times the size of a ribosome (Cooper 2000). More than 50 different proteins are found in the nuclear pore complex.

Messenger RNA is transported from the nucleus to the cytoplasm through the nuclear pore after synthesis and processing within the nucleus. Proteins required for nuclear functions (such as transcription factors, histones, DNA polymerases, RNA polymerases, and splicing factors) are manufactured in the cytoplasm and transported into the nucleus. Transport of molecules through the pore can be a passive or an active, energy-dependent process. The passive transfer of molecules through the nuclear pore complex involves only small molecules and proteins less than 50 kDa in size. These molecules diffuse passively through open aqueous channels approximately 9 nanometers (nm) in diameter.

Most proteins and RNAs are transported actively into and out of the nucleus. During active transport, the nuclear pore can open to a diameter of more than 25 nm. Some proteins are recognized and transported because they carry specific signals (exportins) that bind to a receptor on the transport machinery (Nakielny and Dreyfuss 1999).

Pre-mRNAs and mRNAs are associated with a set of at least 20 proteins, forming a **heterogeneous nuclear ribonucleoprotein complex (hnRNP)** throughout their processing in the nucleus and transport to the cytoplasm. At least two of the hnRNP proteins contain nuclear export signals.

Small nuclear RNAs (snRNAs) function within the nucleus as components of the RNA processing machinery. These RNAs are initially transported from the nucleus to the cytoplasm where they associate with proteins to form functional small nuclear ribonucleoproteins (snRNPs), and then return to the nucleus. Proteins that bind to the 5′ caps of snRNAs appear to be involved in the export of the snRNAs to the cytoplasm, while other sequences are responsible for their transport from the cytoplasm into the nucleus (Cooper 2000).

2.11. Transport of Proteins within the Cytoplasm

Proteins are transferred into the endoplasmic reticulum while they are being translated on membrane-bound ribosomes. The **endoplasmic reticulum (ER)** is a network of membrane-enclosed tubules and sacs (cisternae) that extend from the nuclear membrane throughout the cytoplasm. The entire endoplasmic reticulum is enclosed by a continuous membrane and is the largest organelle of most eukaryotic cells. Proteins destined to remain in the **cytosol** (the fluid portion of the cytoplasm, excluding organelles) or to be incorporated into the nucleus or mitochondria are synthesized on free ribosomes and released into the cytosol when their translation is complete.

Two types of ER, the rough and the smooth, perform different cellular functions. The rough ER is covered by ribosomes on its outer surface and functions in protein processing. The smooth ER functions in lipid metabolism (Cooper 2000). The details of the ER transport are an active area of research, but appear to involve different pathways for tRNA, ribosomal RNAs, and mRNAs (Stutz and Rosbash 1998).

Proteins move across the ER membrane through a hydrophilic channel that is evolutionarily ancient (Matlack et al. 1998). The ER membrane channel translocates proteins and also integrates membrane proteins into the lipid bilayer. Thus, the ER membrane channel must identify the signal sequences of the proteins, open in response to the signal, transport the protein from one side of the membrane to the other, and close. For membrane-spanning proteins, some parts must be moved across the membrane, while others must be left in the cytoplasm. The endoplasmic reticulum has several mechanisms to ensure that only properly folded proteins enter the secretory pathway (Reddy and Corley 1998).

Proteins received from the ER are further processed and sorted for transport to their eventual destinations in the **Golgi complex** (Rothman 1994, Cooper 2000). The Golgi complex is a cytoplasmic organelle specialized for processing and sorting proteins and lipids prior to their transport to lysosomes, the plasma membrane, or secretion. The Golgi is made of flattened membrane-enclosed sacs (cisternae) and associated vesicles. Proteins from the ER enter one side of the Golgi and exit from the opposite side. As the proteins pass through the Golgi they are modified; modifications that occur within the Golgi include glycosylation.

Glycosylation involves adding carbohydrates to proteins. Proteins, lipids, and polysaccharides are transported from the Golgi complex to their final destinations via different kinds of transport vesicles. The precise mechanism by which the Golgi functions remains disputed (Featherstone 1998), but the Golgi appears to be in constant flux, each cisterna emerging from the ER with its load of proteins and then carrying the proteins across the Golgi, while at the same time putting the finishing touches on the proteins.

2.12. mRNA Stability

Many of the genes involved in development are expressed for a short time and in precisely defined domains of the body. These genes often are involved in activating expression of genes that are "downstream" in the developmental pathway. Thus, gene expression must be precisely regulated if cell identity is to be specified. Modifying the rate of transcription of DNA is not always sufficient to maintain fine-tuned developmental processes. Sometimes existing mRNAs and proteins must be removed or deactivated, which involves regulating **mRNA stability** (Surdej et al. 1994).

Several mechanisms are involved in regulating mRNA stability, including removal of the poly(A) tail, premature termination of translation due to a premature termination codon, and, perhaps, mRNA localization. For example, during development of *Drosophila*, the process of cellularization of the blastoderm can result in the destabilization of the *bicoid* mRNA.

2.13. Chaperones and the Proteosome

Maintaining the function of proteins within cells depends on more than transcription and translation. The initial folding of proteins and the assembly of multiprotein complexes sometimes require the assistance of molecular **chaperones**—proteins that catalyze protein folding. Chaperones prevent proteins from aggregating into insoluble, nonfunctional blobs and help them reach and maintain a stable functional state. Understanding of protein folding has increased dramatically in the past few years (Radford 2000). The new view of folding indicates there is not a single, specific folding pathway; rather there are potentially a plethora of routes to the folded protein, and which pathways are used will depend on the amino acid sequence, the topology of the protein, and the experimental conditions. Different routes might be used and/or different intermediates and transition states observed as a consequence of relatively small alterations of a common free-energy profile.

After initial folding and assembly, proteins may be damaged. Such proteins can be rescued (refolded) by chaperones or destroyed by proteases. The efficiency and cost of protein quality depend on a balance among folding, refolding, and degradation (Wickner et al. 1999).

Controlled degradation of proteins inside the cell is essential. For example, protein degradation is important in cell-cycle control, heat shock response, programmed cell death, muscle atrophy, immune response, metamorphosis, development, and differentiation (Mykles 1999). Proteins targeted for destruction are marked by the attachment of a small protein called **ubiquitin**, and, following unfolding, the protein is degraded within a large protein complex known as the **proteosome** (Stuart and Jones 1997). Protein degradation is potentially hazardous to the cell and must be restricted to specific sites and times in

order to prevent the improper destruction of useful proteins. The proteosome provides a compartment to confine the proteolytic action to proteins that carry a degradation signal.

2.14. RNA Silencing or Interference

RNA silencing (RNA interference) appears to be a mechanism for defending against the invasion of mobile DNA elements (transposable elements), which can cause mutations when they insert themselves into or close to a gene.

When double-stranded (ds) RNA is injected into eukaryotic cells by transposable elements and viruses, the ds RNA appears to function as a signal that the cell is being invaded because most RNA in a cell is single-stranded. Thus ds RNA elicits a defense response called RNA silencing/interference (Matzke et al. 2001).

RNA silencing/interference involves cutting the ds RNA up into smaller chunks of about 22 nt. These fragments are then degraded in a second reaction by RNase. The two-step reaction is efficient because each molecule of ds RNA primes several RNase molecules, so the cell can mount a large response to only a few ds RNA molecules (Baulcombe 2001).

RNA silencing has been used experimentally to knock out gene function in order to study the role of specific genes in development. RNA silencing also has been proposed as a method for modifying pest arthropods for pest management programs (see Chapter 14).

2.15. Gene Regulation in Eukaryotes

At any one time, only about 15% of all the genes in an insect cell are turned on (Harshman and James 1998). Thus, insect development, behavior, and reproduction are determined by the expression of different genes at different times in different tissues. Research on the control of gene expression in *Drosophila* is expected to help resolve the fundamental principles of eukaryotic gene regulation (Harshman and James 1998).

Some eukaryotic genes code for essential metabolic enzymes or cell structural components and are expressed constitutively at a specific level in all cells. Such genes are often called **housekeeping genes**, but most genes are not expressed continuously in eukaryotic organisms. After cells differentiate, gene regulation may be influenced by environmental cues such as hormones, nutrients, or temperature. The control of gene expression is called **gene regulation**.

Gene regulation in eukaryotes is less well understood than it is in prokaryotes, but appears to be achieved by:

1. regulating the level of transcription,
2. alternative splicing of mRNA transcripts,
3. DNA amplification,
4. programmed rearrangements of DNA,
5. methylation of cytosine bases, or
6. translational control.

Different genes are regulated differently, and gene regulatory mechanisms are often surprisingly complex, employing more than one method.

Regulation of transcription is the most common method of gene regulation in eukaryotes. Messenger RNA molecules generally are short-lived, probably persisting only a few minutes

or hours. Rapid turnover means that the amount of a particular mRNA in the cell can be controlled by adjusting the rate of transcription of specific genes. Both negative and positive transcriptional regulation can occur.

Transcriptional regulation involves **transcriptional activator proteins** that bind with an upstream DNA sequence to prepare a gene for transcription. They may help assemble a transcriptional complex, or they may initiate transcription by an already-assembled transcriptional complex. Some transcriptional activator proteins have a helix–turn–helix structure, which is a sequence of amino acids that form a pair of alpha helices separated by a bend. These helices fit into the grooves of a double-stranded DNA molecule and allow the proteins to bind to the DNA, although the specificity of the binding is determined by other parts of the protein (Harrison 1991). Examples of helix–turn–helix DNA-binding proteins in insects include the **homeo domain**, which is found in genes such as *engrailed* and *Antennapedia*, which are important in regulating development of *Drosophila melanogaster* (described in more detail in Chapter 3).

Zinc finger proteins are a second type of transcriptional activator protein. They are characterized by loops (fingers) of repeating amino acid sequences each associated with a zinc atom. Zinc finger proteins bind in the major groove of the DNA helix (Figure 2.8) of an upstream DNA sequence to prepare a gene for transcription.

Leucine zippers are a third type of DNA-binding protein. Leucine zippers are DNA binding proteins that contain four to five leucine residues separated from each other by six amino acids. The leucines on two protein molecules can interdigitate and dimerize in a specific interaction with a DNA recognition sequence (Abel and Maniatis 1989).

Hormones may turn on the transcription of specific sets of genes. For example, steroid hormones penetrate a target cell through diffusion because steroids pass freely through the cell and nuclear membranes. The nuclei of target cells contain specific receptor proteins that form complexes with the hormone, which then undergoes modification in its three-dimensional form and enables the receptor–hormone complex to bind with particular sequences in the DNA and stimulate or repress transcription. In the lepidopteran *Manduca sexta*, **ecdysone**, a steroid, acts directly on the genome both to activate and repress genes. Ecdysone initiates and coordinates the molting process and thus the sequential expression of stage-specific genes (Riddiford et al. 1990).

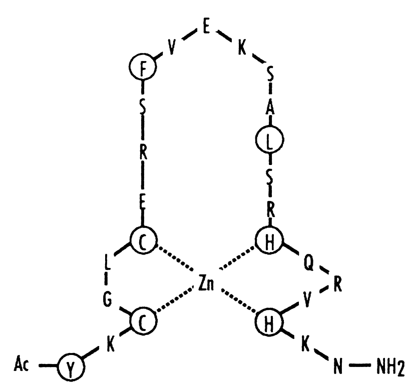


Figure 2.8. Amino acid sequence of a *portion* of a protein containing zinc fingers drawn to illustrate the finger motif. Zinc finger proteins are able to recognize specific DNA sequences. The fingers can intercalate into the DNA molecule and are important in gene regulation.

Hormone–receptor complexes and transcriptional activator proteins bind with specific DNA sequences called enhancers. Enhancer sequences can be found in a variety of sites in relation to the target gene (Figure 2.3). Enhancers can be long or short distances upstream (5′) from the target gene; they may be included in introns within the coding region of the target gene, or even at the 3′ end of the gene. Some enhancers respond to molecules produced inside the cell during development, and many genes are under the control of several enhancers so they can respond to a variety of internal and external molecular signals.

In the living cell, the DNA of the chromosome is tightly bound up with proteins called histones and other proteins that can make transcription of the genes difficult by preventing the association of proteins with the DNA that are needed for gene transcription (Grunstein 1997, Pennisi 1997, Workman and Kingston 1998). At least four different enzymes, called nuclear histone acetylating enzymes, are associated with the transcription complex and add simple chemical groups known as acetyls to the histones to open the DNA up to transcription. In addition, there are five more enzymes that can undo the reaction and remove acetyls from the histones, thus making transcription more difficult.

The physical location of a gene within the nucleus may be important in gene regulation (Cockell and Gasser 1999). For example, correlations have been found between gene silencing and the gene's proximity to a heterochromatic region or to the periphery of the nucleus. Unfortunately, it is not known whether this is a causal relationship; it is possible that gene silencing changes the gene's structure so that it becomes heterochromatic and located near the periphery of the nucleus (Singer and Green 1997). Heterochromatic DNA is usually condensed (making it stain intensely), and any genes within heterochromatic regions usually are inactive.

The components in a typical Class II eukaryotic gene are illustrated in Figure 2.3. A transcriptional complex binds to the promoter to initiate RNA synthesis. The coding regions of the gene (the exons) are interrupted by introns that are eliminated in RNA processing. Transcription is regulated by enhancer elements (numbered 1–6) that respond to different molecules. Because enhancers may respond to different signals or cell conditions, genes can be regulated by a combination of different enhancers. Combinatorial control of gene transcription makes it possible to increase the complexity of gene regulation. If transcription is determined by which pattern of binding state occurs, then a small number of regulatory molecules can yield a large number of different regulatory patterns for different life stages or tissues.

Some genes have two or more promoters. Each promoter usually is active in different cell types, which allows for independent regulation of transcription (Figure 2.9). The different promoters yield different primary transcripts that code for the same polypeptide. Thus, transcription in insect larvae could be controlled by one promoter (Figure 2.9B) and transcription in adults by another (Figure 2.9C).

Different cell types can produce different quantities of a protein or different proteins by **alternative splicing** of mRNA. Differential splicing of exons A and B can result in different rates of synthesis, although the proteins may be identical. In other cases, the proteins produced are different as a result of alternative splicing. Alternative splicing is involved in many aspects of development, including determining the sex of an insect (see Chapter 10). Alternative splicing also makes it possible for protein diversity to be greater than the number of genes in a genome (Graveley 2001).

RNA editing involves altering the sequence of the RNA after it has been transcribed and before it has been translated. RNA editing apparently is widespread in both prokaryotes and

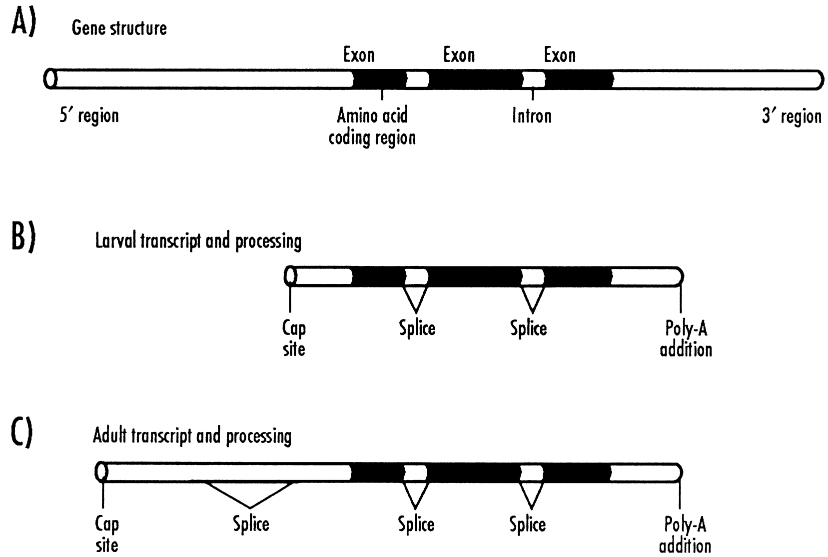


Figure 2.9. Gene regulation can be achieved with the use of alternative promoters. (A) In this gene there are two introns within the coding region. (B) The mRNA transcript in larvae uses the promoter nearest the 5′ end of the coding region. (C) In adults, the promoter further upstream is used, and much of the leader sequence used in larvae is removed by splicing.

eukaryotes and results in functionally distinct proteins produced from a single gene (Maas and Rich 2000).

Gene regulation also can occur by controlling translation of the message into protein. **Translational regulation** can occur in several ways: 1) inability to translate an mRNA molecule unless a particular signal is present, 2) regulation of the longevity of a particular mRNA molecule, 3) regulation of the rate of protein synthesis, and 4) localization of transcripts where translation is required (Lipshitz and Smibert 2000).

An example of inability to translate a mRNA unless a particular signal is present is found in unfertilized eggs, which are biologically static. After fertilization, many new proteins are synthesized, including the mitotic apparatus and cell membranes. However, unfertilized eggs can store large quantities of mRNA for months in an inactive form that abruptly and rapidly becomes active within minutes after fertilization. The timing of translation is thus regulated.

2.16. Insulators and Boundaries

Each gene is embedded within a chromosomal environment of other DNA sequences that have the potential to affect its expression (Bell et al. 2001, Cai and Shen 2001). For example, regulatory elements (enhancers or silencers) associated with nearby genes could disrupt normal expression of a gene. To combat the intrusion of extraneous regulatory elements upon a specific gene, specialized DNA sequences called **insulators** provide a barrier against influences from surrounding DNA sequences. Insulators regulate gene activity by blocking enhancer–promoter interactions when positioned between the enhancer and promoter (Cai and Shen 2001).

Insulators also have the ability to protect against "position effects." When genes are moved from their normal site in the chromosome, they may be expressed differently in the new location. This is often the case with **transgenes** (foreign genes artificially introduced into an organism) (Geyer 1997). This variability in expression may be due to the nearness of an enhancer or silencer or the presence of nearby inactive heterochromatin.

2.17. Chromosome or Gene Imprinting by Methylation in Insects?

Imprinting is a reversible, differential marking of genes or chromosomes that is determined by the sex of the parent from whom the genetic material is inherited. One method of imprinting DNA in some organisms involves methylation. Imprinted genes or chromosomes behave differently.

Methylation of cytosines at the carbon 5 position of CpG dinucleotides is common in many prokaryotes and eukaryotes (Colot and Rossignol 1999, Ng and Bird 1999). In prokaryotes, methylation is apparently part of a defense system against invading DNA parasites. However, in eukaryotes methylation can be associated with several functions, including inhibiting transcription initiation, arresting transcript elongation, serving as a signal for imprinting, and suppressing homologous recombination (Colot and Rossignol 1999). Usually, methylated DNA is inactive or expressed at a very low level. In mammals, DNA methylation is common, and mammalian DNA contains about 1 to 2% of the cytosine as 5-methylcytosine. In mammals, DNA methylation regulates chromatin structure, gene repression, parental imprinting, and X-chromosome inactivation in females. However, DNA methylation is not always the cause of gene inactivity.

The role of methylation in insects is controversial. In some insects, such as scale insects or mealybugs, differential marking of paternally transmitted chromosomes leads to the inactivation or elimination of the paternal chromosomes (White 1973). For example, citrus mealybug males and females develop from fertilized eggs, and there are no sex chromosomes. However, in embryos destined to be males, one haploid set of chromosomes becomes heterochromatic and remains so in most of the tissues. Thus, males are functionally haploid because one chromosomal set, invariably the chromosomes inherited from the father, is inactivated. Bongiorni et al. (1999) found that the chromosomes derived from the citrus mealybug fathers actually are **hypomethylated** (containing a lower amount of methylation of the DNA) than are the chromosomes from the females. As a result, the mechanism by which imprinting and heterochromatinization occurs is unknown in these mealybugs, but methylation of the DNA is not involved (Buglia et al. 1999).

The amount of DNA methylation observed in insects varies. DNA methylation has been found in homopterans, including *Megoura viciae, Planococcus lilacius, Pseudococcus calceolariae, P. obscurus*, and *Myzus persicae*; orthopterans, including *Locusta migratoria, Eyprepocnemis ploras, Pyrgomorpha conica, Gryllotalpa fossor*, and *Baetica ustalata*; and the lepidopteran *Bombyx mori* (Manicardi et al. 1994, Regev et al. 1998). Differential expression of alleles due to imprinting has been observed in *Drosophila melanogaster* (Golic et al. 1998, Regev et al. 1998), but the subject of DNA methylation in *D. melanogaster* has been controversial (Gowher et al. 2000, Lyko et al. 2000). Gowher et al. (2000) estimated that *D. melanogaster* DNA has approximately 50 times less methylcytosine than do mammals, which explains why it is difficult to detect methylation

in *Drosophila*. However, using a sensitive detection method, 5-methylcytosine was found in all stages of *D. melanogaster* development (Gowher et al. 2000). DNA methylation appears to be lacking in some other Diptera (*D. virilis, Sciara coprophila, Musca domestica, Sarcophaga bullata*) or present at only low levels (*Culex bitaeniorhynchus, Chironomus plumosus, Anopheles maculipennis, Aedes albopictus*) (Regev et al. 1998, Tweedie et al. 1999).

2.18. Eukaryotic Genomes and Evolution

The discovery of split genes and RNA splicing has been a critically important finding and has elicited considerable thought regarding the origin and evolution of eukaryotic genomes. Gene regulation, and especially RNA splicing, is probably central to understanding the development of complex multicellular eukaryotic organisms. Alternative RNA splicing produces multiple mRNAs that encode different proteins (Sharp 1994). The spliceosome process for excising introns is probably as old as the ribosomal process for translation. Thus, the eukaryotic cell has two compartments: the nucleus, where the spliceosome processes pre-mRNAs by RNA catalysis; and the cytoplasm, where the ribosome translates mRNAs by RNA catalysis.

Evolution in eukaryotes by changes in RNA processing is aided by the availability of an intron-rich, repeat-rich genome, which allows new gene products to evolve based on changes in RNA processing (Herbert and Rich 1999). Thus, the eukaryotic genome can be thought of as a "junkyard" in which solutions to any number of problems can occur with the assembly of old components into new combinations by changes in RNA processing.

Unlike prokaryotes, eukaryotic organisms make RNA that differs from the DNA in the genome. This allows genomic information to be influenced by information derived from the environment. For example, numerous gene products can act on an mRNA, some affecting nucleotide sequence while others can change the half-life and translatability. Thus, the way in which pre-mRNA is processed in eukaryotes has led to new evolutionary opportunities (Herbert and Rich 1999). A full understanding of the evolution of eukaryotes appears to require knowledge of the continual interplay between RNA and DNA.

The organization of DNA in nuclear chromosomes and in cytoplasmic mitochondria is described in Chapter 3.

General References

Beebee, T., and J. Burke (1988). *Gene Structure and Transcription*. IRL Press, London.

Brown, T. A. (1989). *Genetics, A Molecular Approach*. Chapman and Hall, London.

Cold Spring Harbor Symposia on Quantitative Biology (1998). *Mechanisms of Transcription*. Vol. LXIII. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cooper, G. M. (2000). *The Cell. A Molecular Approach*. 2nd Ed., Am. Soc. Microbiol. Press, Washington, DC.

Garrett, R. A., S. R. Douthwate, A. T. Matheson, P. B. Moore and H. F. Noller, Eds. (2000). *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*. Am. Soc. Microbiol. Press, Washington, DC.

Gesteland, R. F., Ed. (1999). *The RNA World*. 2nd Ed., Monograph 37. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hartl, D. L. (1991). *Basic Genetics*. 2nd Ed. Jones and Bartlett Publ., Boston.

Simons, R. W., and M. Grunberg-Manago, Eds. (1998). *RNA Structure and Function*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

References Cited

Abel, T., and T. Maniatis (1989). Gene regulation. Action of leucine zippers. *Nature* **341**: 24–25.

Baker, D. (2000). A surprising simplicity to protein folding. *Nature* **405**: 39–42.

Baulcombe, D. (2001). Diced defense. *Nature* **409**: 295–296.

Bell, A. C., and G. Felsenfeld (1999). Stopped at the border: boundaries and insulators. *Curr. Opin. Genet. Dev*. **9**: 191–198.

Bell, A. C., A. G. West, and G. Felsenfeld (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**: 447–495.

Blackwood, E. M., and J. T. Kadonaga (1998). Going the distance: a current view of enhancer action. *Science* **281**: 60–63.

Bongiorni, S., O. Cintio, and G. Prantera (1999). The relationship between DNA methylation and chromosome imprinting in the coccid *Planococcus citri*. *Genetics* **151**: 1471–1478.

Buglia, G., V. Predazzi, and M. Ferraro (1999). Cytosine methylation is not involved in the heterochromatization of the paternal genome of mealybug *Planococcus citri*. *Chrom. Res.* **7**: 71–73.

Cai, H. N., and P. Shen (2001). Effects of *cis* arrangement of chromatin insulators on enhancer-blocking activity. *Science* **291**: 493–495.

Cech, T. R. (2000). The ribosome is a ribozyme. *Science* **289**: 878.

Cockell, M., and S. M. Gasser (1999). Nuclear compartments and gene regulation. *Curr. Opin. Genet. Dev.* **9**: 199–205.

Colot, V., and J. L. Rossignol (1999). Eukaryotic DNA methylation as an evolutionary device. *BioEssays* **21**: 402–411.

Cooper, G. M. (2000). *The Cell*: *A Molecular Approach*. Am. Soc. Microbiol. Press, Washington, DC.

Cramer, P., D. A. Bushnell, J. Fu, A. L. Gnatt, B. Maier-Davis, N. E. Thompson, R. R. Burgess, A. M. Edwards, P. R. David, and R. D. Kornberg (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**: 640–649.

Denton, M., and C. Marshall (2001). Laws of form revisited. *Nature* **410**: 417.

Dever, T. E. (1999). Translation initiation: adept at adapting. *Trends Biochem. Sci.* **24**: 398–403.

Featherstone, C. (1998). Coming to grips with the Golgi. *Science* **282**: 2172–2174.

Geyer, P. K. (1997). The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.* **7**: 242–248.

Golic, K. G., M. M. Golic, and S. Pimpinelli (1998). Imprinted control of gene activity in *Drosophila. Curr. Biol.* **8**: 1273–1276.

Gowher, H., O. Leismann, and A. Jeltsch (2000). DNA of *Drosophila melanogaster* contains 5-methylcytosine. *EMBO J.* **19**: 6918–6923.

Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**: 100–107.

Grunstein, M. (1997). Histone acetylation in chromatin structure and transcription. *Nature* **389**: 349–352.

Guo, M., and S. M. Mount (1995). Localization of sequences required for size-specific splicing of a small *Drosophila* intron *in vitro. J. Mol. Biol.* **253**: 426–437.

Harrison, S. C. (1991). A structural taxonomy of DNA-binding domains. *Nature* **353**: 715–719.

Harshman, L. G., and A. A. James (1998). Differential gene expression in insects: transcriptional control. *Annu. Rev. Entomol.* **43**: 671–700.

Hastings, M. L., and A. R. Krainer (2001). Pre-mRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* **13**: 302–309.

Herbert, A., and A. Rich (1999). RNA processing and the evolution of eukaryotes. *Nat. Genet.* **21**: 265–269.

Hilleren, P., and R. Parker (1999). Mechanisms of mRNA surveillance in eukaryotes. *Annu. Rev. Genet.* **33**: 229–260.

Hoffman, A., T. Oelgeschlager, and R. G. Roeder (1997). Considerations of transcriptional control mechanisms: do TFIID–core promoter complexes recapitulate nucleosome-like functions? *Proc. Natl. Acad. Sci. USA* **94**: 8928–8935.

John, S., and J. L. Workman (1998). Just the facts of chromatin transcription. *Science* **282**: 1836–1837.

Kiseleva, E. V. (1989). Secretory protein synthesis in *Chironomus* salivary gland cells is not coupled with protein translocation across endoplasmic reticulum membranes. *FEBS Lett.* **257**: 251–253.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**: 187–208.

Kyrpides, N. C., and C. R. Woese (1998). Universally conserved translation initiation factors. *Proc. Natl. Acad. Sci. USA* **95**: 224–228.

Lee, T. I., and R. A. Young (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**: 77–137.
Lipshitz, H. D., and C. A. Smibert (2000). Mechanisms of RNA localization and translational regulation. *Curr. Opin. Genet. Dev.* **10**: 476–488.
Lyko, F., B. H. Ramasahoye, and R. Jaenisch (2000). DNA methylation in *Drosophila melanogaster*. *Nature* **408**: 538–539.
Maas, S., and A. Rich (2000). Changing the genetic information through RNA editing. *BioEssays* **22**: 790–802.
Manicardi, G. C., D. Bizzaro, P. Assoni, and U. Bianchi (1994). Cytological and electrophoretic analysis of DNA methylation in the holocentric chromosomes of *Megoura viciae* (Homoptera, Aphididae). *Genome* **37**: 625–630.
Maquat, L. E., and G. G. Carmichael (2001). Quality control of mRNA function. *Cell* **104**: 173–176.
Matlack, K. E. S., W. Mothes, and T. A. Rapoport (1998). Protein translocation: tunnel vision. *Cell* **92**: 381–390.
Matzke, M., A. J. M. Matzke, and J. M. Kooter (2001). RNA: Guiding gene silencing. *Science* **293**: 1080–1083.
Mykles, D. L. (1999). Structure and function of arthropod proteasomes. *Mol. Biol. Rep.* **26**: 103–111.
Nakielny, S., and G. Dreyfuss (1999). Transport of proteins and RNAs in and out of the nucleus. *Cell* **99**: 677–690.
Ng, H. H., and A. Bird (1999). DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.* **9**: 158–163.
Nissen, P., J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920–930.
Pennisi, E. (1997). Opening the way to gene activity. *Science* **275**: 155–157.
Preiss, T., and M. W. Hentze (1999). From factors to mechanisms: translation and translational control in eukaryotes. *Curr. Opin. Genet. Dev.* **9**: 515–521.
Radford, S. E. (2000). Protein folding: progress made and promises ahead. *Trends Biochem. Sci.* **25**: 611–618.
Reddy, P. S., and R. B. Corley (1998). Assembly, sorting and exit of oligomeric proteins from the endoplasmic reticulum. *BioEssays* **20**: 546–554.
Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**: 340–345.
Regev, A., M. J. Lamb, and E. Jablonka (1998). The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol. Biol. Evol.* **15**: 880–891.
Riddiford, L. M., S. R. Palli, and K. Hiruma (1990). Hormonal control of sequential gene expression in *Manduca epidermis*. pp. 226–231. In: *Progress in Comparative Endocrinology*, A. Epple, and C. G. Scanes, Eds., Wiley-Liss.
Rothman, J. E. (1994). Mechanisms of intracellular protein transport. *Nature* **372**: 55–63.
Sharp, P. A. (1994). Split genes and RNA splicing. Nobel Lecture. *Cell* **77**: 805–815.
Singer, R. H., and M. R. Green (1997). Compartmentalization of eukaryotic gene expression: causes and effects. *Cell* **91**: 291–294.
Stuart, D. I., and E. Y. Jones (1997). Cutting complexity down to size. *Nature* **386**: 437–438.
Stutz, F., and M. Rosbash (1998). Nuclear RNA export. *Genes Dev.* **12**: 3303–3319.
Surdej, P., A. Riedl, and M. Jacobs-Lorena (1994). Regulation of mRNA stability in development. *Annu. Rev. Genet.* **28**: 263–282.
Tweedie, S., H. H. Ng., A. L. Barlow, B. M. Turer, B. Hendrich, and A. Bird (1999). Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nat. Genet.* **23**: 389–390.
White, M. J. D. (1973). *Animal Cytology and Evolution*. 3rd Ed. Cambridge Univ. Press, London.
Wickner, S., M. R. Maurizi, and S. Gottesman (1999). Posttranslational quality control: folding, refolding and degrading proteins. *Science* **286**: 1888–1893.
Workman, J. L., and R. E. Kingston (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**: 545–579.

3

Nuclear and Extranuclear DNA in Insects

3.1. Overview

DNA is found in chromosomes, which are very complex structures in the nucleus of insect cells. Each chromosome contains a single linear DNA molecule combined with a variety

of proteins, including histones. DNA and histones form structures called nucleosomes. Nucleosomes are arranged in a higher level of organization that serves to condense the DNA. Chromosomes visible by light microscopy may have discrete structures called centromeres, to which spindle fibers attach so that chromosomes are distributed to the daughter cells in an orderly fashion during mitosis and meiosis, although some insect chromosomes have diffuse centromeres along the entire chromosome. The centromeres of eukaryotic chromosomes are complex regions that play a fundamental role in chromosome movement. Chromosomes always have telomeres at the ends, which are specialized structures that help maintain the ends of chromosomes in a stable state. In polytene tissues such as *Drosophila* salivary glands, banding patterns are visible after staining by light microscopy. The bands staining lightly are called euchromatin, and darkly staining bands are called heterochromatin. Eu- and heterochromatin represent active and less active regions, respectively, during somatic interphase. In most insects, there are two copies of each chromosome and hence two copies of each gene (2n = diploid complement), although polyteny or gene amplification can increase the copy number of chromosomes or genes. The reproductive cells (eggs, sperm) contain a single copy (haploid complement = n) of chromosomes.

Chromosomes contain single-copy DNA (DNA present only once in the genome that codes for a polypeptide), highly repetitive DNA, moderately repetitive DNA, intergenic regions, centromeres, and telomeres. Intergenic regions may contain transcription and regulatory information, but a large amount of this region is of unknown function. Many types of transposable elements have been identified in both the coding and noncoding regions of arthropod chromosomes. Transposable elements are DNA sequences capable of moving within and among chromosomes. Transposable elements, or their defective derivatives, make up a significant portion of the middle-repetitive DNA in insect chromosomes.

Arthropods also contain chromosomal DNA in mitochondria, which are located in the cytoplasm of the cell. Mitochondria are considered remnants of a microbial endosymbiont.

3.2. Introduction

DNA in insects is organized into chromosomes, which are complex structures. Different insect species have a different number of haploid chromosomes, ranging from one to 221. Insect chromosomes were among the first chromosomes to be investigated, and studies of insect chromosomes have provided fundamental advances in genetics, including the initial proof that genes are on chromosomes and that spindle fibers exist in living cells and are not fixation artifacts (Ault 1996).

The complete sequence of the *Drosophila* genome, only the third eukaryotic genome to be sequenced after yeast (*Saccharomyces cerevisiae*) and the nematode (*Caenorhabditis elegans*), became available in 2000 (Adams et al. 2000, Celniker 2000). The *Drosophila* genome is about 180 megabases (Mb), and contains approximately 13,600 genes and about 60 Mb of noncoding DNA called heterochromatin (Celniker 2000). Because heterochromatin cannot be cloned stably, the sequences obtained primarily are from the euchromatin regions that contain active genes. The function of the noncoding, heterochromatic DNA is of great interest because it could provide clues about genome evolution and gene regulation.

3.3. C Value Paradox

There appears to be a great deal more DNA in eukaryotic organisms than is actually needed to code for the number of genes estimated for a specific species. This discrepancy is known as the **C value paradox**.

Genome size seems to bear little relationship to organismal complexity or the number of genes encoded. For example, genome size varies widely *among* insect species, with up to 250-fold differences in C values known (Petrov et al. 2000). The locust *Schistocerca gregaria* has a C value of 9,300,000 kilobases (kb), 52-fold more than *Drosophila melanogaster*, but is unlikely to have 52 times as many genes (Wagner et al. 1993). Among 37 species of tenebrionid beetles, nuclear DNA content varies by a factor of 5 (Juan and Petitpierre 1991).

Genome size also can vary *within* species. For example, diploid cells in the mosquito *Aedes albopictus* contain 0.18 to 6 picograms (pg) of DNA, and C values vary by a factor of 3 (from 0.62 to 1.6 pg) among different populations of *A. albopictus* (Kumar and Rai 1990). The amount of DNA in insect cells is difficult to measure because many tissues are polyploid, with different tissues having different degrees of ploidy.

Polyploidy occurs when the amount of DNA in an organism increases over the usual diploid (2n) amount, usually by duplicating the number of chromosomes, perhaps to 3n or 4n or more. Polyploidy can occur throughout an organism's cells or in just some tissues. A few insects are polyploid in all tissues (Otto and Whitton 2000), but many insects have polyploid tissues within a diploid body. For example, the diploid blood cells of *Bombyx mori* contain 1 pg of DNA/blood cell, but a polyploid silk gland cell in the same insect contains 170,000 pg of DNA.

DNA content within cells also varies with developmental stage. At metamorphosis, the amount of DNA in *B. mori* declines by 81% after adults emerge from the pupal stage, which is probably due to histolysis of the polyploid larval silk glands and other polyploid cells.

Noncoding DNA can constitute 30% to more than 90% of the insect genome. This noncoding DNA has been called **junk**, **parasitic**, or **selfish**. There are several hypotheses to explain its persistence in genomes. One suggests that the noncoding DNA performs essential functions, such as global regulation of gene expression. According to this hypothesis, the junk DNA is functional and deletions of such DNA would have a deleterious effect. A second hypothesis is that the noncoding DNA is useless, but is maintained because it is linked physically to functional genes; the excess DNA is not eliminated because it does not affect fitness of the organism and can be maintained indefinitely in the population. A third hypothesis suggests that the noncoding DNA is a functionless parasite that accumulates and is actively maintained by selection. A fourth hypothesis is that the DNA has a structural function, perhaps for compartmentalizing genes within the nucleus, or for maintaining a structural organization (nucleoskeleton) within the nucleus (Manuelidis 1990, Manuelidis and Cher 1990). Of course, all these hypotheses could be correct.

The lack of correlation between genome size and complexity or gene number (C value paradox) remains a topic of study because, unless the noncoding DNA has a function, such DNA constitutes a "load" upon the insect and should be lost over evolutionary time. Petrov et al. (1996) provided evidence that nonessential DNA is lost at a higher rate in *Drosophila* species than in mammalian species, suggesting that differences in genome size may result from persistent differences between organisms in the rate of loss of nonessential DNA. Petrov et al. (2000) provided additional support for this hypothesis by comparing DNA loss in two insect genera (*Laupala* crickets and *Drosophila*) with different genome sizes.

The crickets have a genome size an order of magnitude larger than that of *Drosophila* and eliminate nonessential DNA one-fortieth as quickly.

3.4. Repetitive DNA Is Common in Insects

Much of the noncoding, heterochromatic junk DNA in insects is **repetitive DNA**—specific nucleotide sequences that are repeated several times to millions of times. Repetitive DNA has been classified as **highly repetitive** (sequences repeated several hundred to several million times per genome), or as **moderately repetitive**. Highly repetitive DNA is found in and near centromeres, telomeres, and other heterochromatic regions. Moderately repetitive DNA sequences are repeated 100 to 10,000 times and include genes that code for ribosomal RNA (rRNA) and transfer RNA (tRNA). Moderately repetitive sequences are found in euchromatic regions, as well as in heterochromatic regions of the genome.

Species vary in the number of repeated elements in their genome. For example, *Drosophila melanogaster* has about 30% of its genome as repetitive DNA, but about 60% of the genome of *D. nasutoides* is repetitive DNA. More than 90% of the genome can be noncoding repetitive DNA. Some insects such as aphids have small amounts of repetitive DNA, which could be associated with a faster development time (Ma et al. 1992).

Repetitive and single-copy DNA are present in two different patterns in insect genomes. The **short period interspersion pattern** has single-copy DNA, 1000 to 2000 bp long, alternating with short (200 to 600 bp) and moderately long (1000 to 4000 bp) repetitive sequences. The house fly *Musca domestica*, the Australian sheep blowfly *Lucilia cuprina*, and the wild silk moth *Antheraea pernyi* have this pattern.

Long period interspersion patterns have long (>5600 bp) repeats alternating with very long (>12 kb) uninterrupted stretches of unique DNA sequences. This pattern is often found in species with small genomes (0.1 to 0.5 pg of DNA/haploid genome), including *D. melanogaster*, as well as in the aphid *Schizaphis graminum* (Ma et al. 1992), the midge *Chironomus tentans* (Wells et al. 1976), the fleshfly *Sarcophaga bullata* (Samols and Swift 1979), the honey bee *Apis mellifera* (Crain et al. 1976), and the flour beetle *Tribolium castaneum* (Brown et al. 1990).

Even within an insect family, genome organization can vary. Total DNA in the genome of four mosquito species varies from 0.186 to 0.899 pg, and the amount of repetitive elements varies from 0.009 to 0.150 pg of foldback DNA (Black and Rai 1988). The mosquito *Anopheles quadrimaculatus* has a long period interspersion type of genome organization, but *Culex pipiens*, *Aedes albopictus*, and *A. triseriatus* have the short period interspersion type. Generally, the amounts of foldback, highly repetitive, and moderately-repetitive DNA increase linearly with genome size in these mosquitoes. Intraspecific variation in the amount of highly repetitive DNA was found in *A. albopictus* colonies and may be due to differences in the number or type of transposable elements. The amounts of repetitive DNA in mosquitoes varies from 20% in *An. quadrimaculatus* to 84% in *A. triseriatus* (Besansky and Collins 1992). Because genome organization of relatively few insect species has been studied, it is difficult to determine the significance of these patterns.

Satellite DNA is a type of highly repetitive DNA that differs sufficiently in its base composition from the majority of DNA in a eukaryotic species that it separates out as one or more distinct bands when DNA is isolated by centrifugation with cesium chloride. Satellite DNA is rich in either A+T or G+C sequences and is found in long tandem arrays within the heterochromatic regions of chromosomes.

3.5. Composition of Insect DNA

Insect DNA base ratios are lower than those found in vertebrates, with guanine + cytosine bases (G+C) making up from 32 to 42% of the DNA, compared to 45% for vertebrates (Berry 1985). If base composition were random, 50% of the DNA would be G+C.

DNA in eukaryotes can occur in different configurations. Most genomic DNA exists in the B-helix form, but other configurations are known, including triplex DNA. In **triplex DNA**, the usual A-T and C-G base pairs of duplex DNA are present, but in addition a pyrimidine strand is bound in the major groove of the helix. DNA sequences potentially can form triplex DNA. Such structures appear to be common, are dispersed at multiple sites throughout the genome, and make up as much as 1% of the genome. Triplex DNA was identified in polytene chromosomes of *Chironomus tentans* and *D. melanogaster*, where it was found in the euchromatic bands (Burkholder et al. 1991). Triplex DNA is thought to play a role in helping to condense chromatin.

3.6. Chromosomes Are DNA plus Proteins

Eukaryotes must organize and package their DNA in a sufficiently condensed form that it can fit into a very small space in the nucleus, yet this packaging must be compatible with the ability to separate the DNA strands and unwind the DNA helix during DNA replication and transcription. Furthermore, the packaging must occur rapidly. Precise and rapid replication of DNA is required in many tissues during mitosis and meiosis. How this is achieved is not yet fully resolved (Koshland and Strunnikov 1996, Nicklas 1997).

Eukaryotic genes are on linear DNA molecules, with each chromosome containing a single long DNA molecule. In addition, each chromosome contains an approximately equal amount of proteins with different functions. The proteins include DNA and RNA polymerases and regulatory proteins associated with the DNA. At least five **histones** are associated with the DNA in structures called **nucleosomes** (Figure 3.1). Histone proteins (called H1, H2A, H2B, H3, and H4) contain approximately 100 to 200 amino acids, of which 20 to 30% are arginine and lysine. As a result, the histones have a positive charge which helps histones bind to DNA. DNA sequences that code for the major histones are very highly conserved among eukaryotic species, meaning that they are nearly unchanged over billions of years. Histones apparently are crucial to maintaining chromosome structure and may be crucial to the effective function of DNA as the genetic code (Jenuwein and Allis 2001).

The chromosomal DNA in a nucleosome (called core DNA) is connected by **linker DNA** to the next nucleosome (Figures 3.1, 3.2B). The **core DNA** is protected from digestion by restriction enzymes, but the linker DNA is vulnerable to these enzymes. The core DNA is wound approximately 1.75 times around a histone octamer, consisting of two molecules each of H2A, H2B, H3, and H4. A single molecule of histone H1 is associated with the linker DNA and is apparently helpful in compacting the nucleosome. Linker DNA lengths vary from 20 to 100 nucleotide (nt) pairs in different species and in different cell types within the same organism. The genetic function of the linker DNA and the reasons for variation in its length are unknown.

Origins of DNA replication in higher eukaryotes are larger and more complex than in prokaryotes (see Chapter 2). However, the underlying molecular processes are conserved (Benbow et al. 1992). Replications of origin will be discussed further in Chapter 4.

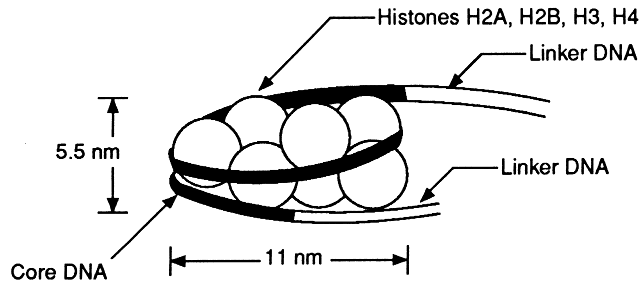


Figure 3.1. DNA is condensed in eukaryotes by being packaged in nucleosomes, which are only the first level of condensation. A nucleosome consists of core DNA wound around two molecules each of the histones H2A, H2B, H3, and H4. These eight histone molecules are called an octamer. Nucleosomes are connected to other nucleosomes by linker DNA. In addition, a single molecule of the histone H1 (not shown) binds in the linker and helps to condense the nucleosome. Nucleosomes are organized into structures called 30-nm fibers. (See text and Fig. 3.2C.)

3.7. Packaging Long, Thin DNA Molecules into Tiny Spaces

Eukaryotes have to solve a serious packaging problem. If eukaryotic chromosomes were simply linear DNA molecules, the average length might be about 5 cm. If these long, thin chromosomes were tangled together inside the nucleus, replication would be difficult and separation of the intertwined chromosomes during mitosis could result in breakage of the chromosomes and subsequent loss of essential genetic information. Therefore, DNA needs to be condensed, yet packaged, so that both DNA replication and transcription can occur without loss or damage. DNA packaging is achieved by a highly organized and hierarchical condensation scheme (Figure 3.2), although the details are not yet resolved completely.

Eukaryotic DNA is **supercoiled**, which means that the double helix is twisted around itself, which begins the condensation process. The next level of compaction is achieved by organizing the DNA into nucleosomes. Nucleosomes occur in a regular pattern, with linker or intervening DNA between each nucleosome (Figure 3.2B). Nucleosomes reduce the length of DNA by a factor of about six to a flexible beaded fiber.

Additional condensation of the DNA occurs when nucleosomes are condensed into a shorter thicker fiber, called the **30-nm fiber** (Figure 3.2C). The structural organization of this 30-nm fiber appears to be an irregular left-handed superhelix with six nucleosomes per turn. It is likely that the 30-nm chromatin fiber form is found in both dividing and nondividing cells (Figure 3.2C). Another level of condensation is thought to occur in metaphase chromosomes (Figure 3.2D–F), although the details of this compaction remain conjectural.

Figures 3.2D, E, and F suggest one model for packing the 30-nm fiber into the highly condensed form found in metaphase chromosomes. Somehow, the length of the chromosomal DNA is reduced by a factor of approximately 10,000 in metaphase chromosomes.

3.8. Structure of the Nucleus

The nucleus contains chromosomes, RNAs, and nuclear proteins in an aqueous solution. It also appears to have an internal structure that organizes the chromosomes and localizes

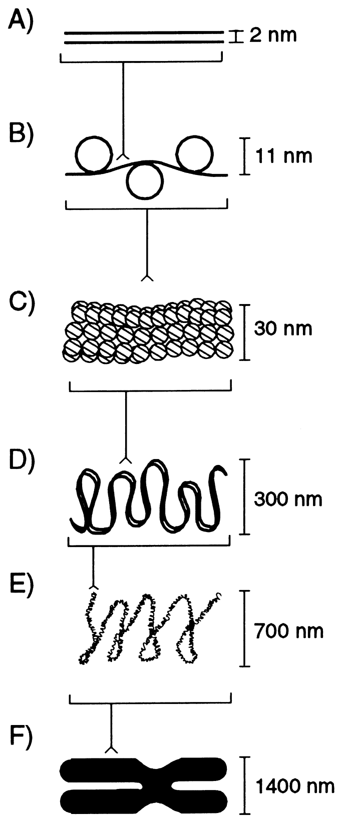


Figure 3.2. DNA packing in eukaryotic chromosomes must be efficient to achieve a dramatic reduction in DNA length. This figure illustrates the method by which DNA is thought to be packed, although the organization of elements in figures D) through F) are controversial. A) naked DNA, B) DNA in nucleosomes, C) 30-nm chromatin fiber, D) the 300-nm fiber made up of looped 30-nm fibers, E) the 700-nm supercoiled structure that comprises the arms of a metaphase chromosome, and F) the metaphase chromosome.

some nuclear functions to specific sites. The most obvious organized region is the nucleolus, which is the site at which the ribosomal RNA (rRNA) genes are transcribed and ribosomal subunits are assembled.

Nucleoli are RNA-rich spherical bodies, not surrounded by a membrane, associated with specific chromosomal segments called the nucleolus organizer. The nucleolus contains multiple copies of tandem arrays of rRNA genes. There are four types of rRNA, the 5S, 5.8S, 18S, and 28S. The 5.8S, 18S, and 28S rRNAs are transcribed as a single unit by RNA polymerase I, yielding a 45S precursor rRNA. The 45S-pre-rRNA is processed into the 18S rRNA of the 40S (small ribosomal subunit) and to the 5.8S and 28S rRNAs of the 60S (large) ribosomal subunit. Transcription of the 5S rRNA, which is found in the 60S subunit, takes place outside the nucleolus and is catalyzed by RNA polymerase III. The nucleolus is particularly important during development, when a large number of rRNA genes are transcribed so that large numbers of ribosomes can be produced.

Much of the **heterochromatin** (chromosome regions that remain condensed during most of the cell cycle and appear to contain mostly inactive genes or noncoding DNA) is localized at the edge of the nucleus, apparently because the heterochromatin binds to a protein of the inner nuclear membrane.

Active euchromatic DNA is arranged in an organized fashion and divided into discrete functional domains that are important in regulating gene expression. Thus, functional euchromatin is nonrandomly distributed within the interphase nucleus. Apparently, each chromosome occupies a discrete region of the nucleus. The chromosomes are closely associated with the nuclear envelope at many sites, with their centromeres and telomeres clustered at opposite poles (Cooper 2000).

3.9. Euchromatin and Heterochromatin

Insect nuclear genomes have two types of chromatin during somatic interphase: euchromatin and heterochromatin. **Euchromatin** is uncoiled during interphase, presumably to allow for gene transcription and condensed during mitosis, with a maximal condensation at metaphase. Euchromatin contains most of the single-copy DNA. In polytene salivary gland chromosomes of *Drosophila*, the darkly staining segments are euchromatic and the intervening (less well stained) regions are heterochromatic.

The term **heterochromatin** was coined originally to define the chromosome regions that remain condensed during most of the cell cycle and have a coiling cycle out of phase with the rest of the genome. Unlike euchromatin, heterochromatin exhibits maximal condensation in nuclei during interphase. Heterochromatin replicates late in the cell cycle, as compared to euchromatin, and may contain a considerable amount of repetitive DNA (Weiler and Wakimoto 1995).

In many organisms, large regions of the chromosome near the centromeres and the telomeres are heterochromatic, and these regions contain primarily repetitive DNA. Heterochromatin is thought to play an essential role in centromere function (Henikoff 2000). It is now thought that heterochromatin is not a type of DNA sequence, but rather a "chromatin state" (Jenuwein and Allis 2001). Potentially, all parts of the genome could enter this state. The repression of transcription in heterochromatin seems to involve a set of proteins and RNA molecules, although the details of how they function remain limited (Hennig 1999, Leach et al. 2000, Redi et al. 2001).

Heterochromatin serves an important role in chromosome mechanics (Wallrath 1998). Without sufficient heterochromatin, chromosomes segregate inappropriately to daughter cells during mitosis. Strangely, a few genes normally located in heterochromatic regions are active there, but become silenced or inactive if moved into euchromatin (Eissenberg and Hilliker 2000).

Heterochromatic regions in *Drosophila melanogaster* can cause **position effect variegation** by inactivating (silencing) euchromatic genes that have been moved to regions adjacent to heterochromatin by chromosomal rearrangements (Wallrath 1998). A change in location of a gene within the nucleus significantly modifies the amount of "gene silencing," perhaps because of its location within the nuclear compartment. Furthermore, foreign genes (transgenes) experimentally inserted into an insect's genome can be silenced because they become heterochromatinized. The original assumption was that the transgenes became inactive because they had been inserted into a heterochromatic site (Henikoff 2000). Thus, understanding the mechanism(s) by which heterochromatin forms will be essential in

improving the function of transgenes inserted into insects developed for pest management programs.

Heterochromatin is hypothesized to serve as a defense mechanism after parasitic DNA invades genomes (Henikoff 2000). In *D. miranda*, heterochromatin forms at clusters of retrotransposons (a type of transposable element) that have recently invaded the genome. The transformation of chromosome regions into heterochromatin might prevent these invasive elements from functioning.

A number of functional genes do occur within heterochromatic regions in *Drosophila*. For example, the Y chromosome of *D. melanogaster* is heterochromatic, yet carries genes that are required for male fertility (Gatti and Pimpinelli 1992). These genes are active in the primary spermatocyte.

3.10. Centromeres

Most chromosomes possess a centromere, which is important in the organization of the developing spindle prior to mitosis or meiosis and the separation of the daughter chromosomes at anaphase. Chromosome fragments lacking centromeres, **acentric fragments**, do not get transmitted to daughter cells, and the genetic information contained on them is eventually lost, which can be lethal.

Some species do not have localized centromeres; rather, the whole chromosome appears to have centromeric properties (**holocentric chromosomes**). If holocentric chromosomes are fragmented, each portion can attach to the spindle and these fragments are not lost at mitosis. Holocentric chromosomes are found in the orders Homoptera, Heteroptera, Mallophaga, Anoplura, and Lepidoptera (White 1973). Even after years of study, centromeres remain poorly understood (Tyler-Smith and Floridia 2001, Henikoff et al. 2001).

Analysis of a centromere in a *Drosophila* minichromosome indicated that the essential core of the centromere is a 220-kb region containing complex DNA. In addition, another 200 kb of DNA on either side is essential to the function of the centromere and contains highly repeated sequences (Murphy and Karpen 1995).

Analysis of a centromere from a standard chromosome confirmed that the *Drosophila* centromere spans 420 kb, more than 85% of which consists of two highly repeated satellite DNAs with the sequences AATAT and AAGAG. The remainder of the centromere consists of interspersed transposable elements, as well as a nonrepetitive segment of AT-rich DNA (Sun et al. 1997). Both the repetitive and nonrepetitive sequences contribute to the centromere function.

The chromosomal region adjacent to the *Drosophila* centromere contains very long blocks of highly repetitive DNA in which simple sequences are repeated thousands of times (satellite DNA). There may be several different satellite DNA types in a given species; for example, three satellites are found near centromeric DNA of *Drosophila virilis*. One has a 5′-ACAAACT-3′ repeat; the second is 5′-ATAAACT-3′; and the third is 5′-ACAAATT-3′. These satellite DNAs apparently are not transcribed, and they may bind proteins essential for centromere function. Likewise, for the sheep blowfly *Lucilia cuprina*, several subfamilies of satellite DNA are present in the centromeric regions of the chromosomes, as well as in the sex chromosomes (Perkins et al. 1992).

Whereas each insect species normally has several types of satellite DNA in the centromere, two parasitic wasps, *Diadromus pulchellus* and *Eupelmus vuilleti*, have only one (Bigot et al. 1990). In these two species, satellite DNA constitutes 15 and 25% of the

genome, respectively. Likewise, approximately 50% of the genome of the tenebrionid mealworm, *Tenebrio molitor*, consists of only one type of satellite DNA which is distributed evenly over the centromeric regions (Plohl et al. 1992). In the tenebrionid *Palorus ratzeburgii*, approximately 31% of the genome is a single type of satellite DNA (Ugarkovic et al. 1992).

3.11. Telomeres

The ends of the chromosomes have distinct structures called **telomeres**. Telomeres have two important functions: 1) maintain the length of chromosomes despite the inability of DNA polymerase to replicate linear DNA ends completely, and 2) distinguish natural chromosome ends from double-stranded breaks in DNA. The latter function, known as capping, is important because damaged DNA with double-stranded breaks is attacked by repair and degradative enzymes. Telomeres also associate with one another, with the nuclear matrix, and with the nuclear envelope, which could be important in maintaining nuclear organization and for meiotic chromosome pairing (Mason and Biessmann 1995).

Molecular analyses indicate telomeres consist of a series of repeated nucleotides and proteins (Blackburn 1991, Wagner et al. 1993, Zakian 1989). As described in Chapter 1, DNA replication conventionally occurs only in the 5′ to 3′ direction and cannot be initiated without a primer, which is usually RNA. After primer removal, gaps would remain at the 5′ ends of new DNA strands in eukaryotes if it were not for telomeres. Telomeres prevent the gradual loss of genetic information from the ends of chromosomes. Recent evidence suggests all eukaryotes use a single-stranded DNA binding protein to cap the telomere (de Lange 2001).

Telomeric DNA sequences and structure appear to be similar among many arthropods, including Hymenoptera, Lepidoptera, Trichoptera, Mecoptera, Coleoptera, Orthoptera, Isoptera, Blattodea, and Crustacea (Okazaki et al. 1993, Sahara et al. 1999). The telomeric sequence, TTAGG, was isolated from the silkworm *Bombyx mori* (Okazaki et al. 1993), where it is repeated over a 6- to 8-kb segment. TTAGG has been found only in telomeres from arthropods and could be ancestral in this group.

Not all insects have the TTAGG sequences in their telomeres, including several dipterans. For example, *Drosophila* has one or more non-LTR retrotransposable elements (called HeT-A and TART) which function as telomeres (Mason and Biessmann 1995). In *Drosophila*, chromosome length is maintained, despite incomplete DNA replication, by the addition of retrotransposon (or retroposon) sequences through transposition, but the Het-A and TART elements never transpose into the gene-rich euchromatic regions. The telomere region of *Drosophila* is much longer and more complex than in standard telomeres. It is not known whether *Drosophila* recruited existing retrotransposon elements to replace the standard telomeres or whether the retrotransposons represent a more ancient arrangement used by the earliest eukaryotes (Pardue and DeBaryshe 1999). Eickbush (1997) suggests that non-LTR retrotransposons gave rise to telomerases and that in early eukaryotes a parasite was recruited by the cell to supply this important function. Perhaps the *D. melanogaster* case can be viewed as a recent example of a similar recruitment event.

Two dipterans, *Chironomus* and *Anopheles*, and some dermapterans, hemipterans, and coleopterans also lack the TTAGG repeat, suggesting that the loss of the TTAGG sequences has occurred independently several times during insect evolution (Sasaki and Fujiwara 2000). In chironomids, a third type of chromosome termination occurs (Kamnert et al. 1997,

Rosen and Edstrom 2000). In *Chironomus pallidivittatus* and *C. tentans*, the size of the repeats is unusually large, consisting of 340 bp and 350 bp, respectively. In *C. pallidivittatus* the telomeric repeat units are present in blocks up to 200 kb.

Telomere terminal transferase, or **telomerase**, was discovered first in the protozoan *Tetrahymena*. Telomerase is a ribonucleoprotein, whose RNA component is essential to its function. Telomerase recognizes single-stranded oligonucleotides ending in -GGG3′ in a variety of eukaryotes and adds GGGG sequences to the ends of chromosomes. Telomerase functions as a kind of reverse transcriptase, because its own RNA codes for a DNA sequence.

Adjacent to the telomeres are subtelomeric repetitive sequences. In most insects, recombination is relatively common in subtelomeric DNA. Subtelomeric DNA is thought to act as a buffer zone to protect nearby genes from DNA loss during DNA replication, to promote the spread of short repeats between telomeres through recombination, to give rise to telomere-to-telomere associations that affect the three-dimensional arrangement of chromosomes in the interphase nucleus, and to promote meiotic pairing of chromosomal homologs (Kamnert et al. 1997).

Telomeres are crucial for the viability of the cell. If telomeres are damaged, progressive loss of DNA occurs and the sticky ends of damaged chromosomes will bind to other chromosomes with sticky ends, resulting in chromosome abnormalities such as dicentric chromosomes (chromosomes with two centromeres) which can lead to chromosome breakage and loss. Our understanding of telomeres continues to increase and could continue to provide some surprises (Greider 1999).

3.12. Chromosomes during Mitosis and Meiosis

Chromosomes are visible by light microscopy during the cell divisions called mitosis and meiosis. The following reviews the basic aspects of these two types of cell division, which are essential for growth, development, and reproduction. Mitosis occurs in somatic cells, whereas meiosis occurs only in germ line cells in eukaryotic organisms.

3.12.1. Mitosis

Somatic cells divide by **mitosis**, which produces two nearly identical daughter cells, each containing the same number of chromosomes as the original cell (Figure 3.3). Prior to the onset of mitosis, the chromosomes within the nuclear membrane are not visible by light microscopy, because they are not condensed. Cells not actively undergoing mitosis are in the **interphase** state.

The **cell cycle** consists of a coordinated set of processes by which a cell replicates all its components and is divided into two nearly identical daughter cells. The coordination of cell growth and periodic chromosome replication and division during the cell cycle has important implications for understanding the evolution of cells, development, and for human medicine (Edgar and Lehner 1996, Novak et al. 1997, Dobie et al. 1999, Zachariae 1999, Zhang 1999).

The cell cycle consists of four phases: $G_1 \rightarrow S \rightarrow G_2 \rightarrow M$. DNA synthesis and chromosome duplication take place during the portion of the cell cycle called the **S phase** (for synthesis), but does not occur during the G_1 and G_2 phases of the cell cycle. The M phase represents **mitosis**, in which the duplicated chromosomes and the cytoplasm are divided into two daughter cells. G_1 is the gap between mitosis and DNA synthesis (S),

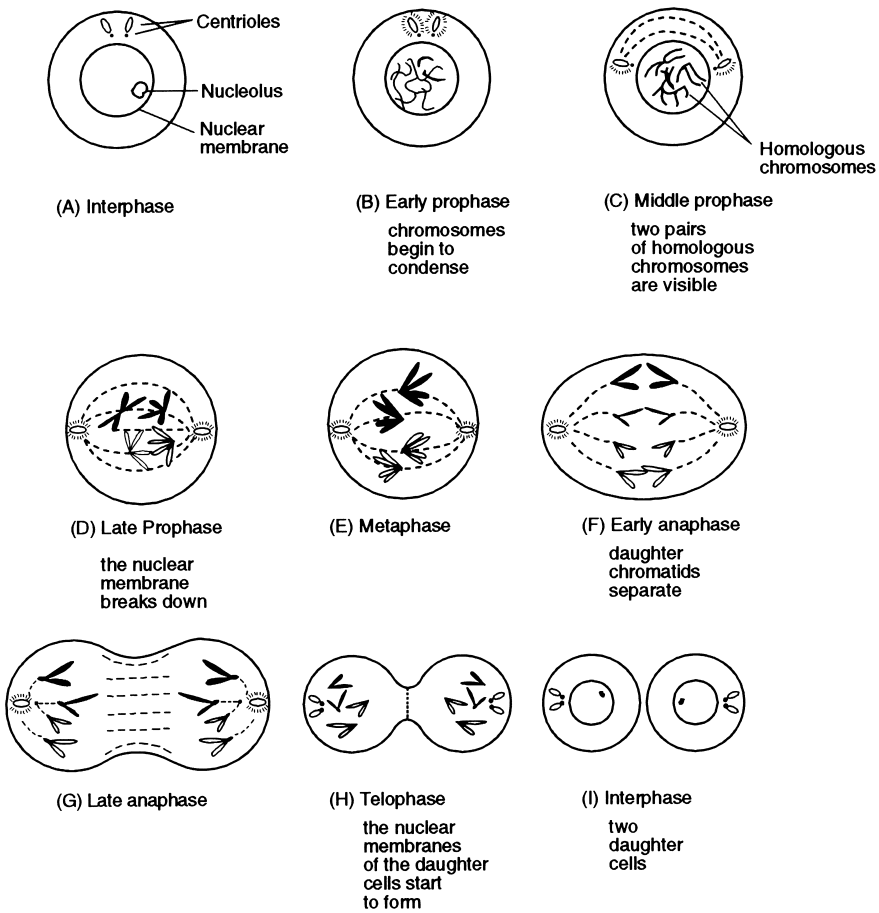


Figure 3.3. Mitosis of a diploid cell involves duplication of each homologous chromosome and its distribution to the daughter cells. DNA replication occurs during interphase. During prophase the two daughter chromatids are attached to each other at the centromere. During metaphase, the chromosomes line up on the metaphase plate, and during anaphase the daughter chromatids separate and begin moving to opposite poles. During telophase the nuclear membranes reform, and two identical daughter cells with a complete complement of chromosomes have been produced.

whereas G_2 is the gap between S and mitosis (M). The length of a cell cycle varies by cell type, but typically lasts approximately 18 to 24 hours, with the process of mitosis requiring 0.5 to 2 hours (Figure 1.14).

Mitosis is divided into four main stages: prophase, metaphase, anaphase, and telophase (Figure 3.3). In **prophase**, the nuclear envelope is still intact and each chromosome condenses to form two visible, thin threads (chromatids) within the nucleus. Because chromosome duplication occurred in the S phase, each chromosome consists of two chromatids connected at the **centromere**. The centromere is the attachment point for the spindle fibers that will draw each of the newly divided chromosomes into their respective nuclei later in mitosis. In late prophase, the nuclear membrane disappears and a mitotic spindle begins to form.

During prometaphase, the spindle develops. The spindle is a complex structure consisting of centrosomes (two centrioles oriented at right angles to one another) and microtubules (hollow protein cylinders consisting of tubulin). The two bundles of fibers extend between the opposite poles of the cell and attach to the centromere of each chromosome (Wolf 1995, Gonzalez et al. 1998). Then, the chromosomes move toward the center of the cell in a plane equidistant from the spindle poles. By the end of **metaphase**, the duplicated chromosomes are lined up on the metaphase plate and are at their most condensed stage, making it easy to examine them for differences in morphology.

During the next stage, **anaphase**, the centromeres divide; the two sister chromatids now have their own centromeres, and so have become independent chromosomes. These newly separated chromosomes move toward the opposite poles. At the end of anaphase, a complete set of chromosomes lies near each opposite pole.

During **telophase**, the chromosomes have reached the spindle poles and the cleavage furrow within the cytoplasm has become visible. The nuclear membrane reforms around each group of chromosomes, the chromosomes decondense, cleavage progresses, and the spindle disappears. The mitochondria often align parallel to the spindle, which may guarantee that they are distributed to both daughter cells. The cytoplasm is divided by a gradually deepening furrow, and a new cell membrane forms. If all has gone well, the result should be the formation of two nearly identical cells with perfectly duplicated genetic information in the nucleus and in the mitochondria within the cytoplasm.

Check points occur during the cell cycle to ensure that the genetic information is duplicated perfectly. During the check points, the genetic material is monitored for integrity and status of replication before the cells commit either to replicate the DNA during S phase, or to segregate it during mitosis (Elledge 1996). If the cell cycle were not well regulated, the cell would be subject to genetic instability or death.

The cell cycle is regulated by protein complexes consisting of cyclins and cyclin-dependent protein kinases (King et al. 1996, Stillman 1996, Piwnica-Worms 1999). The checkpoints involve signal-transduction pathways whose effectors interact with the cyclin/cyclin-dependent protein kinases to block the cell cycle. Blocking the cell cycle allows time for repair of damage at G_1 (before DNA replication) or just before mitosis at the G_2 DNA-damage checkpoint.

Chromosome replication takes place during the S phase, and the duplicated chromosomes remain physically connected (as sister chromatids) until anaphase of mitosis. The cohesion of the sister chromatids is what permits chromosome segregation to take place long after duplication, and this cohesion is due to a multisubunit complex called cohesin. Cleavage of one of cohesin's subunits appears to trigger separation of the sister chromatids at the onset of anaphase (Nasmyth et al. 2000). The ability of eukaryotic cells to delay segregation of the replicated chromosomes until long after they have been duplicated distinguishes the eukaryotic cell cycle from that of bacteria, in which chromosome segregation starts immediately after DNA replication is initiated. The separation of chromosome duplication and segregation has played a central role in the evolution of eukaryotic organisms (Nasmyth et al. 2000). Mitotic chromosome condensation, without which large genomes cannot be partitioned between daughter cells at cell division, would not be possible if chromosome segregation coincided with DNA replication. The gap (G_2) between S and M phases in the cell cycle thus makes possible the evolution of large genomes that can be transmitted safely to daughter cells.

3.12.2. Meiosis

Meiosis probably evolved from a mitosis-like process (van Heemst and Heyting 2000). Meiosis is responsible for two essential aspects of the sexual life cycle in eukaryotes: the transition from the diploid to the haploid state, and the generation of new combinations of alleles. Meiosis occurs only in the germ line (ovaries or testes).

During **meiosis** cells are produced that have a reduced number of chromosomes (the haploid or n number). This means that when the germ cells (eggs and sperm) fuse, the diploid (2n) number of chromosomes is restored (Figure 3.4). If meiosis did not reduce the number of chromosomes to *n*, the number of chromosomes in a sexually reproducing organism would double each generation. Both divisions in meiosis have prophase, metaphase, anaphase, and telophase stages, but their details are different (Figure 3.4). Meiosis may require days or weeks to complete. The essence of meiosis is that only *one* duplication of the chromosomes occurs, but *two* cell divisions occur, producing four haploid gametes from the original diploid cell. Meiosis requires two cell divisions (I and II) to produce daughter cells with the haploid set of chromosomes.

Meiosis I is the *reductional division*, in which the number of chromosomes is reduced from 2n to n. Prophase of meiosis I is a long stage and has been divided into substages (Figure 3.4). During prophase I, the chromosomes condense and become visible. Homologous chromosomes pair and become closely associated along their length. Each homologous chromosome consists of two sister chromatids joined at the centromere; thus the pairing of homologous chromosomes produces a four-stranded structure. During prophase I, the paired chromosomes are able to exchange genetic information by **crossing over**, which results in a shuffling of the genetic information in the gametes. The number of locations where genetic information was exchanged by crossing over often is indicated by the formation of **chiasmata**, which are visible under the microscope during prophase I. Chiasmata result from the physical exchange of nucleotides between chromatids of the homologous chromosomes.

During metaphase I, the two homologous chromosomes are located on opposite sides of the metaphase plate (Figure 3.4). The orientation of each chromosome pair relative to the two poles is random and thus which member of each pair of chromosomes (one set was originally derived from the mother and the other set was originally derived from the father) will move to a particular pole is random. This *random alignment of chromosomes* on the metaphase plate is the basis of Mendel's **Law of Independent Assortment**. Thus, genes originally derived from the individual's mother and father will end up assigned to daughter cells in a random fashion.

During anaphase I, the homologous chromosomes separate from each other and move to opposite poles. This physical separation of homologous chromosomes during anaphase I is the physical basis of Mendel's **Law of Segregation**. After anaphase I, a haploid set of chromosomes consisting of one homolog from each pair is located near each pole of the spindle. During telophase I, the spindle breaks down (Figure 3.4). Chromosomes may pass directly from telophase I to prophase II of meiosis II. Alternatively, there may be a pause between the two meiotic divisions. Chromosome duplication does not occur between meiosis I and II, however.

Meiosis II is similar to a mitotic division, with each daughter cell from meiosis I being replicated, resulting in the production of four haploid cells from the original diploid cell (Figure 3.4). Meiosis II is different from mitosis, however, because the chromatids of a chromosome are usually not identical along their entire length. This is due to the fact that

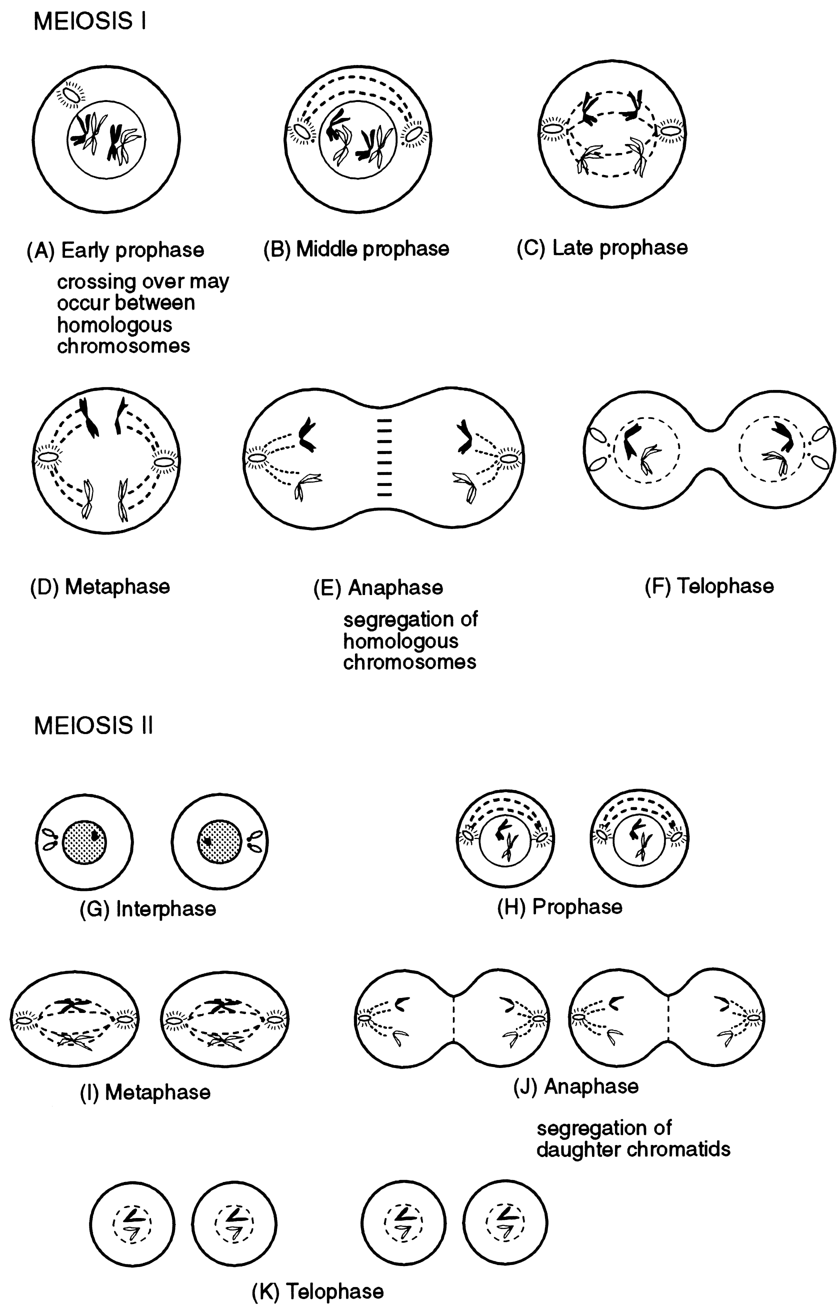


Figure 3.4. Meiosis takes place in the germ line tissues and is a two-step process that results in the production of four haploid cells from a single precursor cell. Meiosis I reduces the chromosome number to the haploid state and involves a lengthy prophase, brief metaphase, anaphase, and telophase. The cells may immediately enter meiosis II. During meiosis II the cells divide to yield four haploid cells.

crossing over could have occurred during prophase of meiosis I and resulted in an exchange of genetic information between the chromatids.

Meiosis has two unusual aspects. One of the most extraordinary aspects of meiosis I is that the two homologous chromosomes that are destined to pair and undergo recombination (crossing over) are able to find each other in a vast set of nonhomologous sequences. How this is achieved is a matter of considerable interest (Roeder 1997, Haber 1998). In *Drosophila*, pairing of homologous chromosomes may be facilitated through specialized pairing sites on the chromosomes. Heterochromatin, especially in the centromeres and telomeres, has been implicated as a mechanism that facilitates chromosomal pairing (Walker and Hawley 2000). During the pairing of homologous chromosomes, an elaborate ladder of protein called the synaptonemal complex is formed that helps to hold them together (Haber 1998).

A second extraordinary aspect of meiosis is the pairing of sister chromatids until their disjunction. This cohesion also is facilitated by protein complexes (van Heemst and Heyting 2000). During pairing, recombination by crossing over occurs at a 100- to 1000-fold higher frequency in meiosis than in mitosis. Recombination tends to occur at certain chromosomal loci called "hotspots" and occurs more often between homologous chromosomes rather than between the sister chromatids.

Crossing over occurs about twice per paired set of chromosomes and serves two roles: the resulting recombination yields new combinations of alleles and plays a mechanical role in separation (disjunction) of the homologous chromosomes at meiosis I (van Heemst and Heyting 2000). Appropriate separation of homologous chromosomes in meiotic anaphase I requires that paired homologous chromosomes, rather than individual chromosomes, line up on the metaphase I spindle. At anaphase I, the homologous chromosomes move to opposite poles (which results in meiosis I being the reductional division).

Each metaphase chromosome has a distinct morphology that is identifiable by staining with lactic–acetic orcein or other stains (Figure 3.5). The location of the centromere allows cytogeneticists to distinguish particular chromosomes. The arms of the chromosome take up stains in a banding pattern that is characteristic of an individual chromosome.

3.13. Chromosome Damage

Chromosome damage probably occurs continuously in all cells. Types of damage range from single base changes, which result from mistakes made by DNA polymerases during replication, to chromosome breakage. Damage is caused by many factors, including the production of metabolic mutagens within the cells. Certain chemicals in the environment, ionizing radiation, and UV light also damage DNA and chromosomes.

Cells have active repair processes to repair such damage. Repairs occur by direct reversal of damage and by excision of a damaged segment of DNA followed by its replacement. Insects no doubt have many genes involved in DNA repair, with some encoding products that recognize DNA damage, some that can excise the damaged region, and others that repair the damage.

Chromosome breaks can occur at any stage of the cell cycle and generally are repaired by rejoining the broken ends, so that the repaired chromosome appears intact. Unfortunately, not all chromosomal damage is repaired, and chromosomal breaks can lead to large-scale rearrangements of chromatin within chromosomes or exchanges of chromatin between nonhomologous chromosomes.

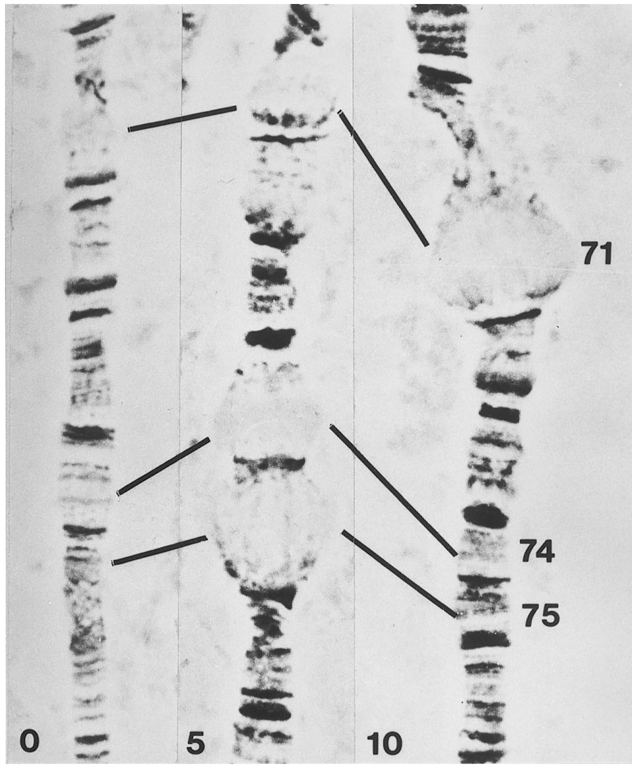


Figure 3.5. Polytene chromosomes from *Drosophila* salivary glands with puffing of different regions.

If the rearranged chromosome lacks a centromere, it is **acentric** and a cell containing the acentric fragment will be unable to transmit this fragment to its daughter cells during meiosis or mitosis, resulting in loss of significant amounts of genetic information, which is usually lethal. Chromosomes that end up with two centromeres (**dicentric**) also are unstable, leading to breaks in the chromosomes if the centromeres are distributed to opposite poles during meiosis or mitosis. This results in breakage and loss of genetic information, which is often lethal.

3.14. Polyteny

In a normal chromosome replication cycle (mitosis), chromosomes condense, replicate, divide, and segregate to daughter cells. In **polytene** cells, 10 or more DNA replication cycles may occur but the daughter chromosomes remain in an extended state and do not separate; such cells become larger and do not divide. The daughter DNA strands stay paired, with homologous regions aligned, which gives rise to a characteristic banding pattern along the length of the chromosome under the light microscope. In some cases, the maternal and paternal homologous chromosomes may synapse, which results in an apparently haploid (n) number of giant chromosomes.

Polyteny is particularly common in larval salivary glands of Diptera, especially in flies from the Drosophilidae, Chironomidae, Cecidomyiidae and Sciaridae, but also occurs in the midgut and fat body in these insects. Polyteny also occurs in Collembola. The number of rounds of DNA strand replication varies from tissue to tissue, with the largest number found in the salivary glands, where there may be as many as 1000 to 2000 chromatids per chromosome.

In *Drosophila* salivary gland chromosomes, the euchromatin regions contain genes while the heterochromatic regions primarily contain repetitive DNA sequences including centromeres and telomeres (Leach et al. 2000). The banding patterns formed by the eu- and heterochromatin make it easy to identify specific sites on *Drosophila* salivary chromosomes (Figure 3.5). About 5000 chromosome bands have been identified in *D. melanogaster*, providing a detailed cytological map. *D. melanogaster* has four pairs of chromosomes; chromosomes two and three are large with central centromeres, and chromosome four is the shortest. Females have two X chromosomes, whereas males have an X and a Y chromosome. The Y chromosome is largely heterochromatic, containing only a few genes.

Because polytene salivary gland chromosomes of *Drosophila* are large and have a well-defined morphology, specific genes can be localized by a procedure called *in situ* hybridization. Radiolabeled DNA or RNA **probes** can be added to salivary gland cells that have been squashed on glass slides. (A probe is a molecule labeled with radioactive isotopes, or another tag, that is used to identify or isolate a specific gene, gene product, or protein.) The labeled probes will anneal to the homologous DNA by base pairing after the chromosomal DNA is denatured (the DNA strands are separated). After any excess probe is washed off, the position of the specific gene can be localized to a specific band or interband region of a specific chromosome by the presence of radioactive grains on an X-ray film. Genes, identified by a particular phenotype, also can be localized to specific sites in polytene salivary gland chromosomes if the mutation is associated with duplications, deletions, inversions, translocations, or other chromosomal abnormalities that can be detected by abnormal banding patterns under the light microscope.

Polytene chromosomes are thought to represent a special case of the more general phenomenon of endopolyploidy (White 1973). In **polyploidy**, an increase in chromosome number occurs within the nuclei of certain tissues without a breakdown in the nuclear membrane. Thus, chromosome duplication takes place, and the chromosomes separate after replication (unlike the situation in polyteny), but no cell division occurs. Many insect cells have 4n, 8n, 16n, and so on, numbers of chromosome sets.

3.15. Chromosomal Puffing

At particular stages in development of many Diptera, some of the genes in salivary gland polytene chromosomes undergo swelling or **puffing**. Puffing is correlated with gene activity. Puffing involves an unraveling of the DNA in a region of the chromosome approximately one to 10 bands in length. The patterns of puffing differ in different instars in *D. melanogaster*, indicating different genes are active in different instars. Puffing is controlled by the hormone ecdysterone, heat shock, and other environmental conditions. The largest puffs contain genes coding for proteins that are produced in very large amounts in the salivary gland, such as the salivary gland secretions and silk. Puffs are associated with extensive transcription of DNA (Figure 3.5).

3.16. B Chromosomes

B chromosomes are a heterogeneous class of chromosomes found in the nucleus and also are called accessory or supernumerary chromosomes. B chromosomes are found in many insects, and they probably originate by several mechanisms, including being derived from autosomes and sex chromosomes in intra- and interspecies crosses (Camacho et al. 2000).

B chromosomes may only be present in some individuals from some populations in a species. B chromosomes have irregular mitotic and meiotic behavior, which allows them to accumulate in the germline, so that they are transmitted at rates higher than those of normal chromosomes.

Over evolutionary time scales, genes on B chromosomes may be silenced, undergo heterochromatinization, and accumulate repetitive DNA and transposons. B-chromosome frequencies in populations result from a balance between their transmission rates and their effects on host fitness. The long-term survival of B chromosomes depends on their ability to survive efforts by their host to eliminate or suppress them because they are often considered to be parasites. Because B chromosomes can interact with standard chromosomes, they could play a positive role in genome evolution if they contribute useful genetic information.

An example of a very interesting B chromosome is that of the PSR (paternal sex-ratio) chromosome of the parasitoid *Nasonia*, which increases in frequency because it is able to destroy paternal chromosomes (described in Chapter 10). The effects of other B chromosomes on their hosts are often unknown.

3.17. Sex Chromosomes

In eukaryotes with identifiable sexes, there generally is a pair of chromosomes called **sex chromosomes**, which are often morphologically different from the rest of the chromosomes (autosomes). In most species, the male is the **heterogametic sex**, which means that it has heteromorphic, only partially homologous sex chromosomes. They are usually called X and Y. The X and Y chromosomes pair in the first prophase of spermatogenesis and as a result of segregation, two types of gametes, one containing the X and one containing the Y chromosome, are produced. Sperm containing the Y that fertilize eggs will result in males whereas sperm containing the X will produce females. Some heterogametic species have males that are XO, lacking a Y chromosome.

Typically, the Y chromosome is smaller than the X and has very few of the genes that are on the X chromosome. The Y is often composed primarily of heterochromatin. The X is usually more like an autosome in function and appearance. However, because the X exists in one copy in the heterogametic sex (XY or XO), some form of dosage compensation is required to equalize the amount of gene product in the two sexes. (Chapter 11 has a discussion of dosage compensation.)

The female is usually the **homogametic sex**, with two X chromosomes, and thus produces only eggs containing an X chromosome. In some insects, such as the Lepidoptera, females are the heterogametic sex. In this case, the sex chromosomes often are designated as W and Z, with the W analogous to the Y of the male (White 1973, Wagner et al. 1993).

3.18. Extranuclear Inheritance in Mitochondrial Genes

Genes located in the nucleus show Mendelian inheritance because they segregate in a regular manner during meiosis. However, not all genes in eukaryotic organisms are located in the nucleus. **Mitochondria** are inherited cytoplasmically and primarily are transmitted through the maternal gametes.

Mitochondria are self-replicating organelles that occur in the cytoplasm of all eukaryotes. Mitochondria are considered to be the descendants of an aerobic eubacterium that became an endosymbiont within an early anaerobic cell that may or may not have contained a nucleus (Kobayashi 1998). The survival of mitochondria within eukaryotic cells has been speculated to have occurred because mitochondria killed their host if the mitochondria were disturbed (Kobayashi 1998).

Mitochondria are thought to have originated once (Gray et al. 1999). Based on studies of DNA sequences, members of the rickettsial subdivision of the α-Proteobacteria, a group of obligate intracellular parasites that include the *Rickettsia*, *Anaplasma*, and *Ehrlichia* are considered to be the closest known relatives of mitochondria (Gray et al. 1999).

Mitochondria have changed during evolutionary time, and the mitochondrial genomes of insects have departed radically from the ancestral pattern (Gray et al. 1999). Mitochondrial genomes range in size from 6 kb to more than 2 Mb (Sogin 1997). Genome size has been reduced following endosymbiosis because some mitochondrial genes became expendable in the internal environment of the host cell (Blanchard and Lynch 2000). Some nuclear genes have replaced the function of the mitochondrial (mt) genes, but much of the reduction in mitochondrial genome size occurred through the *transfer* of mitochondrial protein-coding genes into the nuclear genome. As a result, the mitochondrion has an incomplete set of genes for its own function (Blanchard and Lynch 2000).

The movement of mitochondrial genes into the nuclear genome is an example of **horizontal or lateral gene transfer**. The result of such horizontal transfer means that genes originally located in the mitochondria, but now in the nucleus, must be transcribed, translated, and acquire a sequence that targets the protein produced in the cytoplasm back into the mitochondrion (Ryan et al. 1997). Furthermore, the new nuclear gene must be properly regulated. On reaching the mitochondrion, the protein must be properly folded, modified, and assembled into a larger protein complex. Proteins destined for mitochondria are maintained in their proper form by **molecular chaperones**, proteins that bind to and assist in the folding of proteins into their functional states. Chaperones do not form part of the final protein structure nor do they contain information specifying a particular folding or assembly pathway. Examples of molecular chaperones include proteins produced by the *Hsp70* and *chaperonin* gene families (Ryan et al. 1997). In some insects, including the grasshopper *Podisma pedestris*, mitochondrial pseudogenes (nonfunctional DNA) have been found incorporated in the nuclear genome as well (Bensasson et al. 2000).

Each mitochondrion is surrounded by a double membrane. The inner membrane is highly invaginated, with projections called cristae that are tubular or lamellar. These are the sites of oxidative phosphorylation which result in the formation of **adenosine triphosphate, ATP**, the primary molecule for storing chemical energy in a cell (Saraste 1999). Mitochondrial DNA is a significant component of the total DNA in insect cells. About half of the DNA in an unfertilized *D. melanogaster* egg is mtDNA.

The coexistence of more than one type of mtDNA within a cell or individual, **heteroplasmy**, is thought to be rare in natural populations. Paternal mtDNA either is not transmitted at fertilization or contributes only a small fraction of the mtDNA in the developing embryo, and the paternally derived mitochondria typically are lost during development. Because insect mitochondria are transmitted from mother to progeny, they are inherited asexually (Birky 1995).

The cellular mechanisms that regulate the replication and distribution of mitochondria to daughter cells at each cell division are beginning to be understood (Yaffe 1999). Until recently, it was assumed that inheritance of mitochondria was a passive process, a consequence of their random diffusion throughout the cytoplasm. Now, it is known that mitochondria are associated with the cytoskeleton and move in coordinated ways during cell division and differentiation (Yaffe 1999).

Mitochondria contain distinctive ribosomes, tRNAs, and aminoacyl-tRNA synthetases (Gray 1989, Sogin 1997, Kobayashi 1998). Mitochondria have their own genetic code that differs slightly from the universal genetic code. The mitochondrion of *Drosophila yakuba* codes for 37 genes: 2 are rRNA, 22 are tRNA, and 13 are protein genes that code for subunits of enzymes functioning in electron transport or ATP synthesis (Clary and Wolstenholm 1985, Figure 3.6).

Knowledge of the organization and evolution of insect mitochondrial genomes is being derived from analysis of the complete sequences of mitochondria isolated from an increasing number of species, including *Drosophila yakuba*, *D. melanogaster*, *D. simulans*, *D. mauritiana* and *D. sechellia*, the mosquitoes *Aedes albopictus* and *Anopheles quadrimaculatus*, the grasshopper *Locusta migratoria*, the honey bee *Apis mellifera*, the louse *Heterodoxus macropus*, the silkworm *Bombyx mori*, the blowfly *Phormia regina*, the kissing bug *Triatoma diimidiata*, and the screwworm *Cochliomyia hominivorax* (Clary and Wolstenholm 1985, Cockburn et al. 1990, Goldenthal et al. 1991, Crozier and Crozier 1993, Flook et al. 1995, Lewis et al. 1995, Ballard 2000, Lessinger et al. 2000, Lee et al., GenBank, Shao et al. 2001, Dotson and Beard 2001). Partial DNA sequences of mitochondria have been obtained from many other insects and are deposited in GenBank.

More detailed investigation of insect mitochondrial genomes may provide some contradictions to the generalizations provided. For example, mitochondrial genomes greater than 20 kb have been found in three species of curculionid beetles (*Pissodes strobi*, *P. nemorensis*, and *P. terminalis*) (Boyce et al. 1989). The large size (30 to 36 kb) in these three *Pissodes* species is due to an enlarged A+T enriched region (9 to 13 kb) and a series of 0.8 to 2.0 kb tandemly repeated sequences adjacent to the A+T region. Every weevil sampled in all three species had two to five distinct size classes of mtDNA (exhibited heteroplasmy). The magnitude of the size differences, the number of size classes found within individual weevils, and the abundant mtDNA heteroplasmy are unusual (Boyce et al. 1989).

The dogma that mtDNA is exclusively inherited in a maternal fashion has been questioned in *Drosophila* and marine mussels. Incomplete maternal inheritance of mtDNA occurs in *Drosophila simulans* (Satta et al. 1988, Matsuura et al. 1991), and the high level of heteroplasmy found in the three *Pissodes* species could be due to paternal transmission of mtDNA, although Boyce et al. (1989) did not document paternal transmission actually occurred.

Mitochondrial chromosomes are circular, supercoiled, double-stranded DNA molecules. The mitochondrial chromosome of *Drosophila* contains approximately 18.5 kb of DNA, and each mitochondrion contains multiple copies of the chromosome. Mitochondrial genes in insects lack introns, and intergenic regions usually are small or absent. The ribosomes found in the mitochondria are smaller than the ribosomes in the cytoplasm.

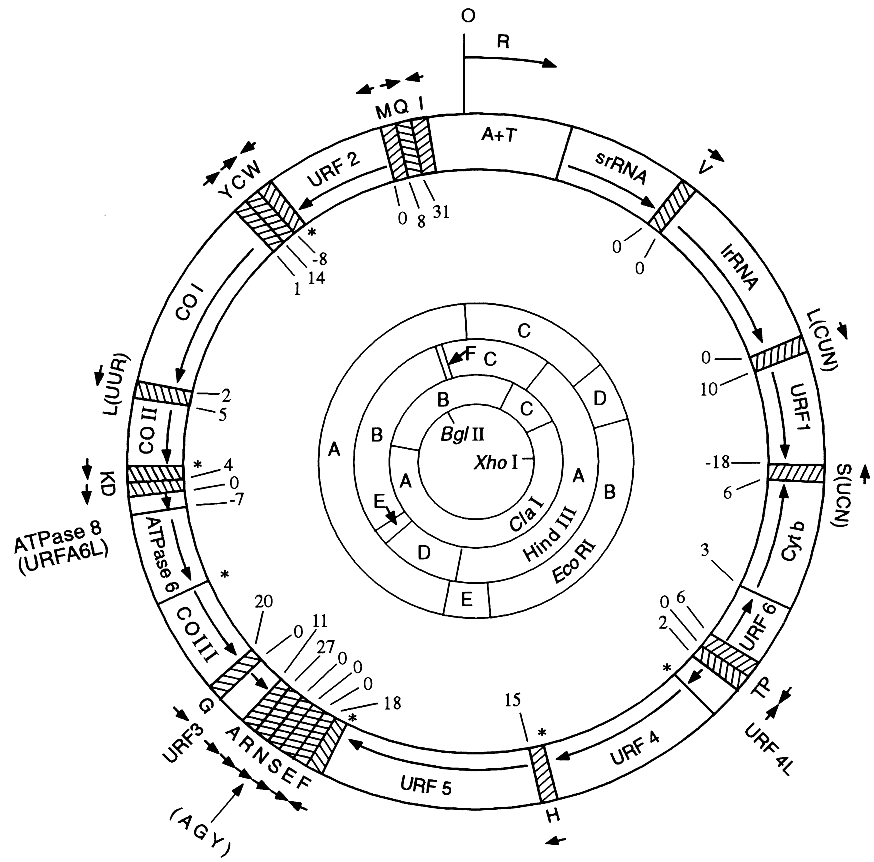


Figure 3.6. Diagram of the circular DNA molecule from the mitochondria of *Drosophila yakuba*. The outside circle shows the open reading frames (URF1 to URF6 and URF 4L) that code for subunits of the respiratory chain NADH dehydrogenase and of the genes coding for cytochrome *b*, cytochrome *c* oxidase subunits I, II, and III, and ATPase subunits 5 and 6. The origin and direction of replication are indicated by O and R. The variable A+T region is shaded. The arrows indicate the direction of gene transcription. The tRNA genes are crosshatched and indicated by their single-letter amino acid codes. lrRNA and srRNA are the large and small rRNA genes. The numbers on the inside of the outer circle are the numbers of apparently noncoding nucleotides that occur between the genes. The innermost circles indicate restriction fragments produced with the enzymes indicated (from Clary and Wolstenholm 1985).

Most eggs and somatic cells contain hundreds or thousands of mtDNA molecules, so a new mutation can result in a situation in which two or more mtDNA genotypes coexist within an individual (heteroplasmy). Heteroplasmy, however, is apparently a transitory state in germ cells. Thus, the majority of individuals are effectively haploid with regard to the number of types of mtDNA transmitted to the next generation.

Mitochondrial DNA evolves faster than single-copy nuclear DNA because mitochondria are relatively inefficient in repairing errors during DNA replication or after DNA damage. In Hawaiian *Drosophila*, mtDNA appears to evolve about three times faster than the genes of nuclear DNA (Moritz et al. 1987). Because mtDNA does not code for proteins involved

directly in its own replication, transcription, or translation, mtDNA has a large number of length mutations and transitions.

Mitochondrial DNA can be amplified easily from mitochondria by the polymerase chain reaction (PCR) (see Chapter 8) because there are multiple copies in each cell. Mitochondria are easier to purify from cells than a specific segment of nuclear DNA. Mitochondria have a specific buoyant density and high copy number within cells. Isolation of mitochondria by centrifugation is relatively easy, making mtDNA a useful subject for systematics or population genetics studies, as will be described in Chapters 12 and 13.

3.19. Transposable Elements Are Ubiquitous Agents That Alter Genomes

Every insect genome probably contains several types of transposable elements (TEs) (Berg and Howe 1989, Craig et al. 2001). An organism may contain active and inactive TEs.

Transposable elements are genetic elements that can move from one chromosomal site to another, and are usually present in multiple copies within a genome. TEs usually consist of a large proportion of the repetitive DNA. The ubiquity of TEs in a diverse array of organisms has raised a number of unanswered questions about their evolutionary impact. New TEs are still invading and spreading within insect populations, and the role of TEs in insect evolution and genetic manipulation will be discussed further in Chapters 4, 8, and 14.

The diversity of arthropods and their genetic systems has only been hinted at. In Chapter 4 we will explore additional details of genome organization, developmental processes, and diversity in insects.

References Cited

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, and P. G. Amanatides, et al. (2000). The genome sequence of *Drosophila melanogaster. Science* **287**: 2185–2195.

Ault, J. G. (1996). Contributions of insect systems to the chromosome theory of inheritance and to our understanding of chromosome behavior on the spindle. *Intern. J. Insect Morphol. Embryol.* **25**: 93–114.

Ballard, J. W. O. (2000). Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Mol. Evol.* **51**: 48–63.

Benbow, R. M., J. Zhao, and D. D. Larson (1992). On the nature of origins of DNA replication in eukaryotes. *BioEssays* **14**: 661–669.

Bensasson D., D. X. Zhang, and G. M. Hewitt (2000). Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol. Biol. Evol.* **17**: 406–415.

Berg, D. E., and M. M. Howe, Eds. (1989). *Mobile DNA*. Am. Soc. Microbiol., Washington, DC.

Berry, S. J. (1985). Insect nucleic acids, pp. 219–253. In: *Insect Physiology Biochemistry and Pharmacology*, Vol. 10, G. A. Kerkut, and L. I. Gilbert, Eds., Pergamon, Oxford.

Besansky, N. J., and F. H. Collins (1992). The mosquito genome: organization, evolution and manipulation. *Parasitology Today* **8**: 186–192.

Bigot, Y., M. Hamelin, and G. Periquet (1990). Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vuilleti* (Hymenoptera). *Mol. Biol. Evol.* **7**: 351–364.

Birky, C. W., Jr. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**: 11331–11338.

Black, W. C., and K. S. Rai (1988). Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. *Genet. Res. Cambridge* **51**: 185–196.

Blackburn, E. H. (1991). Structure and function of telomeres. *Nature* **350**: 569–573.

Blanchard, J. L., and M. Lynch (2000). Organellar genes. Why do they end up in the nucleus? *Trends Genet.* **16**: 315–320.

Boyce, T. M., M. E. Zwick, and C. F. Aquadro (1989). Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* **123**: 825–836.

Brown, S. J., J. K. Henry, W. C. Black, and R. E. Denell (1990). Molecular genetic manipulation of the red flour beetle: genome organization and the cloning of a ribosomal protein gene. *Insect Biochem.* **20**: 185–193.

Burkholder, G. D., L. J. P. Latimer, and J. S. Lee (1991). Immunofluorescent localization of triplex DNA in polytene chromosomes of *Chironomus* and *Drosophila*. *Chromosoma* **101**: 11–18.

Camacho, J. P. M., T. F. Sharbel, and L. W. Beukeboom (2000). B-chromosome evolution. *Phil. Trans. R. Soc. Lond. B* **355**: 163–178.

Celniker, S. E. (2000). The *Drosophila* genome. *Curr. Opin. Genet. Dev.* **10**: 612–616.

Clary, D. O., and D. R. Wolstenholm (1985). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**: 252–271.

Cockburn, A. F., S. E. Mitchell, and J. A. Seawright (1990). Cloning of the mitochondrial genome of *Anopheles quadrimaculatus*. *Arch. Insect Biochem. Physiol.* **14**: 31–36.

Cooper, G. M. (2000). *The Cell: A Molecular Approach*. Am. Soc. Microbiol., Washington, DC.

Craig, N. L., R. Craigie, M. Gellert, and A. M. Lambowitz (2001). *Mobile DNA II*. Am. Soc. Microbiol., Washington, DC.

Crain, W. R., E. H. Davidson, and R. J. Briteen (1976). Contrasting patterns of DNA sequence arrangement in *Apis mellifera* (honeybee) and *Musca domestica* (housefly). *Chromosoma* **59**: 1–12.

Crozier, R. H., and Y. C. Crozier (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**: 97–117.

de Lange, T. (2001). Telomere capping—one strand fits all. *Science* **292**: 1075–1076.

Dobie, K. W., K. L. Hari, K. A. Maggert, and G. H. Karpen (1999). Centromere proteins and chromosome inheritance: a complex affair. *Curr. Opin. Genet. Dev.* **9**: 206–217.

Dotson, E. M., and C. B. Beard (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol. Biol.* **10**: 205–215.

Edgar, B. A., and C. F. Lehner (1996). Developmental control of cell cycle regulators: a fly's perspective. *Science* **274**: 1646–1652.

Eickbush, T. H. (1997). Telomerase and retrotransposons: which came first? *Science* **277**: 911–912.

Eissenberg, J. C., and A. J. Hilliker (2000). Versatility of conviction: heterochromatin as both a repressor and an activator of transcription. *Genetica* **109**: 19–24.

Elledge, S. J. (1996). Cell cycle checkpoints: preventing an identity crisis. *Science* **274**: 1664–1672.

Flook, P. K., C. H. Rowell, and G. Gellissen (1995). The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* **41**: 928–941.

Gatti, M., and S. Pimpinelli (1992). Functional elements in *Drosophila melanogaster* heterochromatin. *Annu. Rev. Genet.* **26**: 239–275.

Goldenthal, M. J., K. A. McKenna, and D. J. Joslyn (1991). Mitochondrial DNA of the blowfly *Phormia regina*: restriction analysis and gene localization. *Biochem. Genet*. **29**: 1–11.

Gonzalez, C., G. Tavosanis, and C. Mollinari (1998). Centrosomes and microtubule organization during *Drosophila* development. *J. Cell Sci.* **111**: 2697–2706.

Gray, M. W. (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**: 25–50.

Gray, M. W., G. Burger, and B. F. Lang (1999). Mitochondrial evolution. *Science* **283**: 1476–1481.

Greider, C. W. (1999). Telomeres do D-loop-T-loop. *Cell* **97**: 419–422.

Haber, J. E. (1998). Searching for a partner. *Science* **279**: 823.

Henikoff, S. (2000). Heterochromatin function in complex genomes. *Biochim. Biophys. Acta* **1470**: 1–8.

Henikoff, S., K. Ahmed, and H. Malik (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**: 1098–1102.

Hennig, W. (1999). Heterochromatin. *Chromosoma* **108**: 1–9.

Jenuwein, T., and C. D. Allis (2001). Translating the histone code. *Science* **293**: 1074–1080.

Juan, C., and E. Petitpierre (1991). Evolution of genome size in darkling beetles (Tenebrionidae, Coleoptera). *Genome* **34**: 169–173.

Kamnert, I., C. C. Lopez, M. Rosen, and J. E. Edstrom (1997). Telomeres terminating with long complex tandem repeats. *Hereditas* **127**: 175–180.

King, R. W., R. J. Deshaies, J. M. Peters, and M. W. Kirschner 1996. How proteolysis drives the cell cycle. *Science* **274**: 1652–1658.

Kobayashi, I. (1998). Selfishness and death: raison d'etre of restriction, recombination and mitochondria. *Trends Genet.* **14**: 368–373.

Koshland, D., and A. Strunnikov (1996). Mitotic chromosome condensation. *Annu. Rev. Cell. Dev. Biol.* **12**: 305–333.

Kumar, A., and K. S. Rai (1990). Intraspecific variation in nuclear DNA content among world populations of a mosquito, *Aedes albopictus* (Skuse). *Theor. Appl. Genet.* **79**: 748–752.

Leach, T. J., H. L. Chotkowski, M. G. Wotring, R. L. Dilwith, and R. L. Glaser (2000). Replication of heterochromatin and structure of polytene chromosomes. *Mol. Cell. Biol.* **20**: 6308–6316.

Lee, J. S., Y. S. Kim, S. H. Hwang, J. S. Lee, D. S. Lee, and D. S. Suh. *Bombyx mori* mitochondrian complete genome. *GenBank* #AF149768.

Lessinger, A. C., A. C. Martins Junqueira, T. A. Lemos, E. L. Kemper, F. R. da Silva, A. L. Vettore, P. Arruda, and A. M. L. Azeredo-Espin (2000). The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* **9**: 521–529.

Lewis, D. L., C. L. Farr, and L. S. Kaguni (1995). *Drosophila melanogaster* mitochondrial DNA, completion of the nucleotide sequence, and evolutionary comparisons. *Insect Mol. Biol.* **4**: 263–278.

Ma, R. Z., W. C. Black IV, and J. C. Reese (1992). Genome size and organization in an aphid (*Schizaphis graminum*). *J. Insect Physiol.* **38**: 161–165.

Manuelidis, L. (1990). A view of interphase chromosomes. *Science* **250**: 1533–1540.

Manuelidis, L., and T. L. Cher (1990). A unified model of eukaryotic chromosomes. *Cytometry* **11**: 8–25.

Mason, J. M., and H. Biessmann (1995). The unusual telomeres of *Drosophila*. *Trends Genet.* **11**: 58–62.

Matsuura, E. T., H. Fukuda, and S. I. Chigusa (1991). Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Reunion. *Genet. Res. Camb.* **57**: 123–126.

Moritz, C., T. E. Dowling, and W. M. Brown (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**: 269–292.

Murphy, T. D., and G. H. Karpen (1995). Localization of centromere function in a *Drosophila* minichromosome. *Cell* **82**: 599–609.

Nasmyth, K., J. M. Peters, and F. Uhlmann (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. *Science* **288**: 1379–1384.

Nicklas, R. B. (1997). How cells get the right chromosomes. *Science* **275**: 632–637.

Novak, B., A. Csikasz-Nagy, B. Gyorffy, K. Nasmyth, and J. J. Tyson (1997). Model scenarios for evolution of the eukaryotic cell cycle. *Phil. Trans. R. Soc. Lond. B* **353**: 2063–2076.

Okazaki, S., K. Tsuchida, H. Maekawa, H. Ishikawa, and H. Fujiwara (1993). Identification of a pentanucleotide telomeric sequence, $(TTAGG)_n$, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.* **13**: 1424–1432.

Otto, S. P., and J. Whitton (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**: 401–437.

Pardue, M. L., and P. G. DeBaryshe (1999). Telomeres and telomerase: more than the end of the line. *Chromosoma* **108**: 73–82.

Perkins, H. D., D. G. Bedo, and A. J. Howells (1992). Characterization and chromosomal distribution of a tandemly repeated DNA sequence from the Australian sheep blowfly, *Lucilia cuprina*. *Chromosoma* **101**: 358–364.

Petrov, D. A., E. R. Lozovskaya, and D. L. Hartl (1996). High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384**: 346–349.

Petrov, D. A., T. A. Sangster, J. S. Johnston, D. L. Hartl, and K. L. Shaw (2000). Evidence for DNA loss as a determinant of genome size. *Science* **287**: 1060–1062.

Piwnica-Worms, H. (1999). Fools rush in. *Nature* **401**: 535–537.

Plohl, M., B. Borstnik, V. Lucijanic-Justic, and D. Ugarkovic (1992). Evidence for random distribution of sequence variants in *Tenebrio molitor* satellite DNA. *Genet. Res. Camb.* **60**: 7–13.

Redi, C. A., S. Garagna, H. Zacharias, M. Zuccotti, and E. Capanna (2001). The other chromatin. *Chromosoma* **110**: 136–147.

Roeder, G. S. (1997). Meiotic chromosomes: it takes two to tango. *Genes Dev.* **11**: 2600–2621.

Rosen, M., and J. E. Edstrom (2000). DNA structures common for chironomid telomeres terminating with complex repeats. *Insect Mol. Biol.* **9**: 341–347.

Ryan, M. T., D. J. Naylor, P. B. Hoj, M. S. Clark, and N. J. Hoogenraad (1997). The role of molecular chaperones in mitochondrial protein import and folding. *Intern. Rev. Cytol.* **174**: 127–193.

Sahara, K., F. Marec, and W. Traut (1999). TTAGG telomeric repeats in chromosomes of some insects and other arthropods. *Chromosome Res.* **7**: 449–460.

Samols, D., and H. Swift (1979). Genomic organization in the flesh fly *Sarcophaga bullata*. *Chromosoma* **75**: 129–153.

Saraste, M. (1999). Oxidative phosphorylation as the fin de siecle. *Science* **283**: 1488–1493.
Sasaki, T., and H. Fujiwara (2000). Detection and distribution patterns of telomerase activity in insects. *Eur. J. Biochem.* **267**: 3025–3031.
Satta, Y., N. Toyohara, C. Ohtaka, Y. Tasuno, T. Watanabe, E. T. Matsuura, S. I. Chigusa, and N. Takahata (1988). Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritiana*. *Genet. Res. Camb.* **52**: 1–6.
Shao, R., N. J. H. Campbell, and S. C. Barker (2001). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Pthiraptera). *Mol. Biol. Evol.* **18**: 858–865.
Sogin, M. L. (1997). History assignment: when was the mitochondrion founded? *Curr. Opin. Genet. Dev.* **7**: 792–799.
Stillman, B. (1996). Cell cycle control of DNA replication. *Science* **274**: 1659–1664.
Sun, X., J. Wahlstrom, and G. Karpen (1997). Molecular structure of a functional *Drosophila* centromere. *Cell* **91**: 1007–1019.
Tyler-Smith, C., and G. Floridia (2000). Many paths to the top of the mountain: diverse evolutionary solutions to centromere structure. *Cell* **102**: 5–8.
Ugarkovic, D., M. Plohl, V. Lucijanic-Justic, and B. Borstnik (1992). Detection of satellite DNA in *Palorus ratzeburgii*: analysis of curvature profiles and comparison with *Tenebrio molitor* satellite DNA. *Biochimie* **74**: 1075–1082.
van Heemst, D., and C. Heyting (2000). Sister chromatid cohesion and recombination during meiosis. *Chromosoma* **109**: 10–26.
Wagner, R. P., M. P. Maguire, and R. L. Stallings (1993). *Chromosomes A Synthesis*. Wiley-Liss, New York.
Walker, M. Y., and R. S. Hawley (2000). Hanging on to your homolog: the roles of pairing, synapsis and recombination in the maintenance of homolog adhesion. *Chromosoma* **109**: 3–9.
Wallrath, L. L. (1998). Unfolding the mysteries of heterochromatin. *Curr. Opin. Genet. Dev.* **8**: 147–153.
Weiler, K. S., and B. T. Wakimoto (1995). Heterochromatin and gene expression in *Drosophila*. *Annu. Rev. Genet.* **29**: 577–605.
Wells, R. H. Royer, and C. P. Hollenberger (1976). Non *Xenopus*-like DNA organization in the *Chironomus tentans* genome. *Mol. Gen. Genet.* **147**: 45–51.
White, M. J. D. (1973). *Animal Cytology and Evolution*. 3rd Ed. Cambridge Univ. Press, Cambridge, UK.
Wolf, K. W. (1995). Spindle membranes and spindle architecture in invertebrates. *Micron* **26**: 69–98.
Yaffe, M. P. (1999). The machinery of mitochondrial inheritance and behavior. *Science* **293**: 1493–1497.
Zachariae, W. (1999). Progression into and out of mitosis. *Curr. Opin. Cell Biol.* **11**: 708–716.
Zakian, V. A. (1989). Structure and function of telomeres. *Annu. Rev. Genet.* **23**: 579–604.
Zhang, P. (1999). The cell cycle and development: redundant roles of cell cycle regulators. *Curr. Opin. Cell Biol.* **11**: 655–662.

4

Genetic Systems, Genome Evolution, and Genetic Control of Embryonic Development in Insects

4.1. Overview

Molecular genetics has revolutionized our understanding of insect gene structure, organization, regulation, and development. One of the greatest surprises has been the discovery that genomes are very dynamic over evolutionary time. For example, large portions of the insect genome may consist of different families of transposable elements (TEs) that can alter gene structure and function, a chromosomal organization, and transfer horizontally between species. The discovery that gene amplification can be involved in resistance to insecticides in aphids and mosquitoes has opened new avenues for understanding this evolutionary and economic problem. Research also suggests that the genetic information contained in microbial symbionts such as *Wolbachia* may play an essential role in speciation and evolution of some insects. Other symbionts may provide essential nutrients or other physiological services. Many insects contain three or four (or more) genomes: nuclear, mitochondrial, one or more gut symbionts, and *Wolbachia*, which questions the concept of the "biological individual."

Insect nuclear genomes are diverse; although many species have diploid males and females, some have haploid males and diploid females (arrhenotoky), and some have only females (thelytoky). In some, diploid males may undergo chromosome heterochromatinization and loss during development to become haploid (parahaploidy). The diversity of genome organization in insects reflects their long evolutionary history.

An understanding of the stages of embryonic development and many of the major genes that influence these stages in *Drosophila melanogaster* is emerging. Evolutionary comparisons of development among organisms are developing into a new field of study called "evo-devo." Three stages of embryonic development occur in *D. melanogaster* that are determined by maternal and zygotic genes: First, the polarity of the embryo is determined, primarily by maternal-effect genes. Next, segmentation genes influence the development of major bands or parasegments. The third determination is accomplished by the interaction of homeotic genes that provide a finer definition of the segmental structures. Analyses of *Drosophila* development may have broad significance for understanding development in all eukaryotic organisms.

4.2. Introduction

Insects are extremely numerous and diverse. Insects make up about half of all described species and about three-fourths of all described animals. There are about 883,475 described insect species in more than 762 families organized in 32 orders (Daly et al. 1998). Many insect species may yet remain undescribed because some estimates suggest there are as many as 2 million insect species. Insects have a long evolutionary history, live in a great variety of habitats, exhibit diverse types of lifestyles, have an extraordinary range of structural variations, eat an astonishing variety of food, and are among the most abundant animals on earth. The long evolutionary history of insects has provided sufficient time for them to develop a diversity of genetic systems (see Chapter 12 for an overview of insect evolution).

This chapter provides only a small sample of the diversity of insect genetic systems. More extensive reviews can be found in White (1973) and Wagner et al. (1993). This chapter also provides an overview of the diversity of microbial symbionts associated with insects, the diversity of their transposable elements, and gene regulation in some insect-specific genes, as well as a brief introduction to the molecular genetics of insect embryonic development.

4.3. Genetic Systems in Insects

Most insects are diploid (2n) in their somatic cells and haploid (n) in their gametes. Other systems can be found; some insect groups are parthenogenetic and may be polyploid, including species in the Orthoptera (Blaberidae, Tettigoniidae), Homoptera (Coccidae Delphacidae), Embioptera (Oligotomidae), Lepidoptera (Psychidae), Diptera (Chamaemyiidae, Chironomidae, Psychodidae, Simuliidae), Coleoptera (Ptinidae, Chrysomelidae, Curculionidae), and Hymenoptera (Diprionidae, Apidae) (Otto and Whitton 2000). Polyploid insects usually are 3n or 4n, but exceptions include curculionid weevil species that are 5n and 6n (Retnakaran and Percy 1985). Parthenogenesis has not been found in the Diplura, Protura, Odonata, Plecoptera, Dermaptera, Grylloblattodea, Zoraptera, Megaloptera, Mecoptera, and Siphonaptera, although only a few species in these groups have been examined carefully.

Parthenogenesis can be divided into three major types: arrhenotoky, thelytoky, and deuterotoky. **Deuterotoky** involves the development of unfertilized eggs into either males or females, and at least one insect, a mayfly, is reported to exhibit facultative deuterotoky (White 1973). In **arrhenotoky**, insects are haplodiploid, with males developing from unfertilized haploid eggs while females develop from fertilized diploid eggs. The entire order Hymenoptera and many species in the Homoptera, Thysanoptera, and Coleoptera are arrhenotokous (Hartl and Brown 1970, White 1973).

Thelytokous insect species have females only. **Thelytoky** has arisen repeatedly in evolution, consists of several types, and can be induced experimentally in a number of ways (White 1973; see Chapter 10 for examples). In some cases of thelytoky, eggs only develop after penetration by a sperm (pseudogamy or gynogenesis), but the sperm nucleus degenerates without fusing with the egg nucleus so that it makes no genetic contribution to the embryo. The sperm may be derived from the testis or ovotestis of a hermaphrodite or from a male of a different, but closely related, species.

Thelytoky may be the sole mode of reproduction in a species, or it may alternate with sexual reproduction in regular manner (cyclical thelytoky), as happens in some aphids (Hales et al. 1997), gall wasps, and some cecidomyiids. In species that reproduce by cyclical thelytoky, genetic recombination is possible, but in species with complete thelytoky there is no

way in which mutations that have occurred in two unrelated individuals can be combined in a third.

Thelytokous reproduction can be induced in the eggs of many species by pricking the egg or exposing it to chemical agents or heat. In a number of normally bisexual insects, a few eggs deposited by virgin females can hatch spontaneously, and the incidence of such egg hatch can be increased by artificial selection. White (1973) suggests that the capacity for artificial parthenogenesis, induced thelytoky, or facultative thelytoky indicates that some capacity for parthenogenesis is probably present in all eggs. Thelytokous species or thelytokous populations of bisexual species have been found in the Diptera, Hymenoptera, Lepidoptera, Orthoptera, and Coleoptera.

In the Homoptera, both arrhenotoky and thelytoky occur (Retnakaran and Percy 1985), but even more complex genetic systems can be found (White 1973, Haig 1993). For example, in some mealybugs (Pseudococcidae), both males and females develop from fertilized eggs but, in the embryos that develop into males, the paternally derived chromosomes become heterochromatic, genetically inactive, and are not transmitted to the male progeny. This genetic system has been called **parahaploidy** (Brown and Nur 1964, Nur 1990). Some method of **chromosome imprinting** is probably involved to ensure that the paternally derived chromosomes are eliminated and not the maternally derived ones. The mechanisms involved in chromosomal imprinting could be associated with methylation of DNA (Sapienza et al. 1987, Solter 1988, Wagner et al. 1993).

4.4. Endopolyploidy Is Common in Somatic Tissues of Arthropods

The discussion of ploidy is confusing because, in most insects, some of the somatic tissues exhibit high levels of endopolyploidy. For example, haploid male honey bees have about the same amount of DNA as females in some of their somatic tissues because nuclei of the male undergo compensatory endomitosis so that equal amounts of DNA are present. In some cases, haploid males are known to exhibit higher levels of endopolyploidy in some tissues than the diploid females of the same species.

4.5. Genetics of Insects Other Than *D. melanogaster*

Much of what we know about the genetics of insects is derived from the study of *Drosophila* species (Ashburner 1989, Brody 1999). Extensive genetic information is available for *D. melanogaster*, including a physical map (Kafatos et al. 1991) and the complete sequences of the genome (Adams et al. 2000, Hawley and Walker 2000, Jabbari and Bernardi 2000, Otto 2000; see also Chapter 6 for additional details on the *Drosophila* Genome Project). For updated information on the *Drosophila* genome and other aspects of *Drosophila* biology, search the Internet at The Interactive Fly and FlyBase, A Database of the *Drosophila* Genome.

Relatively little genetic information is available for the vast majority of the 883,475 known insect species. For example, in 1993, sufficient genetic information was available to develop genetic **linkage maps** for only 27 (Heckel 1993). A linkage map involves identifying specific chromosomes by one or more genetic markers, traditionally by phenotypic mutants. Mutations on sex (X) chromosomes are most easily identified, because they exhibit a characteristic mode of inheritance. Most of the well-studied species

are dipterans: species of *Drosophila*, several mosquitoes, the screwworm *Cochliomyia hominivorax*, the sheep blowfly *Lucilia cuprina*, the Mediterranean fruit fly *Ceratitis capitata*, the Oriental fruit fly *Bactrocera dorsalis*, and the housefly *Musca domestica*. Two coleopterans (*Tribolium castaneum* and *T. confusum*), one orthopteran (the cockroach *Blattela germanica*), hymenopterans (*Habrobracon juglandis*, *Nasonia vitripennis*, *Apis mellifera*) and one lepidopteran (*Bombyx mori*) make up the rest (Wright and Pal 1967, Robinson 1971, King 1975, Sokoloff 1966, 1977, Steiner et al. 1982, Heckel 1993, Severson et al. 1993, 2001). Additional genetic studies of the honey bee *Apis mellifera*, the Mediterranean flour moth *Ephestia kuehniella*, *Heliconius* butterflies, the butterflies *Papilio glaucus* and *Colias eurytheme*, and the tobacco budworm *Heliothis virescens* were reviewed by Heckel (1993). The tenebrionid *Latheticus oryzae*, the fruit fly *Rhagoletis pomonella*, and the grasshoppers *Melanoplus sanguinipes* and *Locusta migratoria* have also been studied genetically (Chapco 1983, Sokoloff 1966). Genetic studies of honey bees and silk moths have potential for improving the management of these beneficial insects (Rinderer 1986, Robinson 1971, Tazima 1964, Tazima et al. 1975).

Analyzing genes, development, and genetic systems from insects other than *Drosophila melanogaster* could help solve both basic and applied problems. *D. melanogaster* may be a highly specialized insect with unique genetic characteristics. A detailed genetic map of the X chromosome of the malaria vector *Anopheles gambiae* was produced using microsatellite DNA markers (Zheng et al. 1993), and plans are underway to sequence the entire 260 million base pairs of *A. gambiae* (Anonymous 2001, Balter 2001, Severson et al. 2001). Studies of the genomes of other economically important insects such as the Mediterranean fruit fly, the silk moth, the flour beetle *T. castaneum*, the mosquito *A. aegypti*, and the honey bee are under way (Brown et al. 1990, Warren and Crampton 1991, Crozier and Crozier 1993, Zheng et al. 1991, 1993, 1996, Besansky and Powell 1992, Hunt and Page 1995, Shi et al. 1995, Beye et al. 1998, Beeman and Brown 1999, Rai and Black 1999, Wu et al. 1999, Tan et al. 2001, Yasukochi 1998).

4.6. Dynamic Insect Genomes

Until recently, the eukaryote genome was considered to be relatively stable, with every cell having the same DNA sequences in the same amounts and in the same location. Genomes were perceived to respond slowly to evolutionary pressures. It is now apparent that somatic genomes are more diverse than previously imagined, with polyteny, polyploidy, and gene amplification occurring in different tissues at different developmental stages in the organism (Edgar and Orr-Weaver 2001). It is also clear that DNA can move within the nuclear genome via a wide array of transposable elements.

DNA has been found in interesting structures outside the nuclear chromosomes and mitochondrion, but their significance is unresolved. For example, covalently closed circular DNAs that appear to be derived from chromosomal DNA have been found in cell cultures of *Drosophila* (Gaubatz 1990). Much of this circular DNA is repetitive chromosomal DNA and may be associated with gene amplification during development or DNA rearrangements during aging (Gaubatz 1990). Some circular DNA molecules in *D. melanogaster* embryos apparently contain 5S ribosomal RNA genes, satellite DNA, or histone genes (DeGroote et al. 1989).

Minichromosomes have been found in *D. melanogaster* that apparently originated from the transposable element TE1 (Block et al. 1990). The minichromosome contains two

structural genes, *white* and *roughest*, from the *Drosophila* X chromosome and part of chromosome 2. This minichromosome was relatively stable and inherited by 33 to 47% of progeny, which indicates that it contains a centromere. Centromere-like elements lacking chromosome arms have been found in the phorid *Megaselia scalaris* (Wolf et al. 1991). The function of these elements is unknown, but they could be B chromosomes that have been reduced to a minimal size.

4.7. B Chromosomes

These are a heterogeneous class of often heterochromatic chromosomes, sometimes referred to as accessory or supernumerary chromosomes that occur in plants and animals. B chromosomes may have little effect on the phenotype, may differ in number from one cell type to another, and may occur only in some individuals of the species. B chromosomes may not segregate normally in mitosis and meiosis (Wagner et al. 1993). A B chromosome in the parasitic wasp *Nasonia vitripennis* causes the compaction and loss of paternally derived chromosomes in fertilized eggs, leading to the production of all male progeny in this arrhenotokous (haplodiploid) species (Eickbush et al. 1992). Thus, the notion that insect genomes simply consist of nuclear and mitochondrial chromosomes should be discarded (Pardue 1991).

During embryonic development of some insects, special germ-line-limited chromosomes are eliminated from those cells that will become somatic cells. The loss occurs because these chromosomes lag during early cleavage divisions. Occasionally, however, these supernumerary chromosomes have been found in the somatic cells of the chironomid *Acricotopus lucidus* over many generations (Staiber 1987).

The notion that eukaryotic genes should contain introns is not always sustained. For example, hemoglobin genes sequenced from the midge *Chironomus thummi* lack introns, even though they show sequence homology with vertebrate hemoglobin genes which do contain introns (Antoine and Niessing 1984). Because the cloned *Chironomus* genes were expressed *in vivo*, the hypothesis that they are pseudogenes was rejected. (A **pseudogene** is a gene with a close resemblance to a known gene, but it is nonfunctional because mutations prevent normal transcription or translation.) An alternative explanation is that the hemoglobin genes originated by **reverse transcription** of spliced mRNA in germ-line cells. Reverse transcription involves synthesis of DNA from a messenger RNA template, which lacks introns, to produce cDNA. If this intronless cDNA subsequently became integrated in the *C. thummi* genome, then the hemoglobin gene would lack introns.

Several major categories of nuclear DNA are known: unique-sequence, middle-repetitive, and highly repetitive. Within the middle-repetitive class of DNA, examples will be presented of some particularly interesting insect genes.

4.8. Unique-Sequence DNA in the Nucleus

Most genetic information is contained in unique-sequence DNA. The proportion of unique sequences varies among species. For example, among four Lepidoptera, *Antheraea pernyi*, *Hyalophora cecropia*, *Bombyx mori*, and *Manduca sexta*, the proportions of unique DNA range from 55 to 80% (Berry 1985).

Some unique-sequence DNA is present in multiple copies in specific insect cells or tissues. This occurs by one of two mechanisms: multiple copies of unique sequences can

occur if the cells are polyploid (polyploidy means that cells contain multiple copies of each chromosome, with $n>2$), or multiple copies of unique sequences also can occur through **gene amplification** in which a portion of the chromosome is replicated. For example, the chorion genes of *Drosophila* are amplified during specific stages of chorion production (see below), although this amplification is limited to ovarian follicle cells. It now appears that some insects that are resistant to pesticides have amplified esterase genes (see below).

4.9. Middle-Repetitive DNA in the Nucleus

Middle-repetitive DNA is found in more than one copy, but still in modest amounts. Such sequences include genes that code for ribosomal RNAs (rRNA), transfer RNAs (tRNA), histones, transposable elements (TEs), and developmentally regulated multigene families such as actins, cuticle genes, heat shock genes, larval serum genes, silk genes, and yolk protein genes. One solution to producing large amounts of gene product in a relatively short time and in a coordinated manner is to duplicate the gene. Duplicated genes may be present in tandem arrays on the same chromosome or may be present on separate chromosomes.

4.9.1. Heat Shock Genes

The heat shock response originally was discovered in *D. melanogaster* and has since been found in organisms ranging from bacteria to man. Heat shock genes are activated in response to environmental stresses such as heat or chemical shock. The heat shock proteins are present in small amounts in many cells in the absence of stress, but rapidly increase after stress. Heat shock genes are an evolutionarily conserved response to stress in all organisms (Morimoto et al. 1992).

If *Drosophila* are exposed to a severe heat shock (about 40°C), most die. If they undergo a mild shock at 33°C, additional heat shock proteins are synthesized, and many flies then can survive subsequent heat shocks at 40°C. In *D. melanogaster* nine chromosomal sites puff in response to heat shock, and specific mRNAs are produced that code for seven heat shock proteins. There are several types (or families), including the *hsp70*, *hsp83*, and the small heat shock gene family (Pauli et al. 1992).

The *hsp70* gene is virtually inactive in unstressed cells, but hsp proteins become very abundant during and after heat shock, accounting for 1% of the total cellular protein (Feder and Krebs 1997). There are 10 copies of the *hsp70* gene in *Drosophila*. It is the most abundant and highly conserved. At the amino acid level, the *Drosophila* hsp70 protein shares 73% overall similarity with that of the human and 50% with that of the bacterium *E. coli*. In addition, seven cognate genes of *hsp70* are constitutively expressed and may be important during *Drosophila* development. The hsp70 proteins are molecular chaperones, minimizing aggregation of peptides in nonnative conformation.

The *hsp83* gene products are general chaperones involved in several developmental pathways in *D. melanogaster* and have both housekeeping and stress-related functions (Rutherford and Lindquist 1998, Mayer and Bukau 1999). The *hsp83* gene also appears to be involved in evolutionary changes in developmental processes (Rutherford and Lindquist 1998). When the *D. melanogaster hsp83* gene is mutated or impaired, variability in many adult structures is induced, with specific variants depending on the genetic background. This phenotypic variability is caused by multiple, previously silent, modified gene products that are suppressed by the normal function of the *hsp83* gene in *D. melanogaster*.

Thus, *hsp83* gene products buffer variation, allowing genetic variation to accumulate under neutral conditions. When the organism is stressed by heat or cold, the hidden variants are expressed and selection then could lead to the continued expression of these traits, providing a mechanism for promoting evolutionary change in an otherwise entrenched developmental process. The protein may act as a "capacitor for morphological evolution" (Rutherford and Lindquist 1998).

The small *hsp* gene family includes genes encoding *hsp22*, *hsp23*, *hsp26*, and *hsp27* proteins, which are expressed at several developmental stages.

4.9.2. Histone Genes

The histone gene family codes for the five histones that serve as the basic proteins in eukaryotic chromosomes. The basic unit of chromosomes, the nucleosome, is composed of 146 bp of DNA coiled around the histone octamer, two molecules each of histone H2A, H2B, H3 and H4 (Figure 3.1). Linking two nucleosomes is a small stretch of DNA to which the fifth histone, H1, is bound. The histone genes share regulatory sequences and are coordinately expressed. In some species, there are tissue- or stage-specific gene sets. In *Drosophila*, the histone genes are tandemly repeated and closely linked. The histone genes of the midge *Chironomus thummi* are different from those found in *D. melanogaster* (Hankeln and Schmidt 1991).

Histone genes typically lack introns. It is thought that introns were eliminated because these genes must be expressed efficiently and rapidly during development. Histone proteins could be produced more efficiently if the pre-mRNA did not need to be spliced to remove introns. Having histone genes organized in a tandem repeat structure also ensures that there will be equivalent amounts of the five proteins produced.

There is a 10-fold difference in copy numbers of histone genes in three species of *Drosophila* (*melanogaster*, *hydei*, and *hawaiiensis*) (Fitch et al. 1990). *D. melanogaster* has 100 tandemly arranged histone genes, far more genes than would be required for the maximal rate of transcription during development. By contrast, *D. hydei* has 5, and *D. hawaiiensis* has about 20 tandem histone repeat copies per haploid genome. In *D. melanogaster* the histone genes are located adjacent to a heterochromatic region of chromosome 2, whereas they are located in euchromatic regions in the other two species. This suggests that there are more histone genes in *D. melanogaster* to compensate for the fact that the genes are less active because they are located near heterochromatin.

4.9.3. Immune Response Genes

Insects defend themselves against bacteria, viruses, fungi, and parasitoids with both cellular and humoral immune responses (Gillespie et al. 1997, Khush and Lemaitre 2000, Carton and Nappi 2001). *D. melanogaster* defends against microbial attack by both constitutive and inducible responses. The first line of defense against microbes is structural and comprises the exoskeleton, the peritrophic membrane that lines the gut, and the tracheal linings. In addition, insects typically maintain a low pH and digestive enzymes and antibacterial lysozymes in their midguts.

Infections induce local immune responses that include the synthesis and secretion of peptides in barrier tissues such as the tracheal and gut epithelium. In addition, systemic responses are activated that result in encapsulation of pathogens by blood cells, melanization of parasites and pathogens, phagocytotic uptake of pathogens by blood cells, and the

production by the fat body of antifungal and antibacterial peptides that are secreted into the hemolymph where they accumulate to high concentrations (Khush and Lemaitre 2000). A number of antibacterial proteins and peptides, such as cecropins, attacins, lysozymes, and defensins, are produced by protein families. Genes belonging to one protein family are often tightly clustered, grouped in one area, or located on the same chromosome arms of *D. melanogaster* (Khush and Lemaitre 2000).

4.9.4. Ribosomal Genes

The ribosome is the site in the cell where proteins are synthesized (Frank 2000). The ribosome is a particle made of two subunits, each formed of an intricate mesh of RNAs and proteins. Protein synthesis is a serious business, and it has to be done quickly and accurately. A typical protein takes approximately 15 seconds to make (Frank 2000). Protein folding, which ensures the proper function of the protein, relies on the location of particular amino acids, which can be jeopardized by even a single point mutation. Typically, ribosomes have an error rate of only 1 in 1000 to 10,000 amino acids. Ribosomes take up much of the cell's mass, and much of the cell's metabolism is devoted to making ribosomal proteins and RNAs. Ribosomes interact with mRNAs, initiation factors, and transfer RNAs during protein synthesis; more than 120 macromolecular components are needed to produce polypeptides in ribosomes (Kaulenas 1985).

Different arthropod species have different numbers of ribosomal genes located in the nuclear chromosomes. For example, *Drosophila erecta* has 160 genes while *D. hydei* has more than 500 (Berry 1985). The fungus fly *Sciara coprophila* contains only 65 to 70, one of the lowest numbers reported (Kerrebrock et al. 1989). Most insect genomes have between 200 and 500 rRNA genes. The ribosomal genes of *Drosophila* are arranged in two clusters, one in the nucleolar organizer of each of the sex chromosomes (Williams and Robbins 1992). The 5.8S, 18S, and 28S rRNAs are transcribed as a single unit, which is then processed in the nucleus to provide the separate subunit RNAs. Ribosomal genes comprise 2% of the total genome and about 20% of the middle-repetitive sequences of *D. melanogaster*.

4.9.5. Silk Genes

Silk is used in cocoons by Lepidoptera, to produce an egg stalk by Neuroptera, and to produce underwater prey-capture nets by Trichoptera (Craig 1997). Silks are composed of one or more proteins called **fibroins**, proteins composed of several simple amino acid sequences in reiterated arrays (Craig 1997, Sezutsu and Yukuhiro 2000). The silk gland provides a model system for cell biologists and molecular geneticists to study gene regulation and development. Silk gland cells of *Bombyx mori* are polyploid (up to 20-fold), which may explain how silk moth larvae produce huge amounts of silk proteins within a short period of time (5 to 6 days) prior to pupating. The cells from the posterior silk gland produce fibroin; those from the middle part store fibroin and **sericin**, a mixture of four to six hot-water-soluble polypeptides. Sericin binds strands of raw silk fibers together. Silk proteins have an unusual amino acid composition, with a predominance of glycine and alanine in fibroin, and serine in sericin (Prudhomme et al. 1985).

Although *Drosophila melanogaster* does not produce silk, a silk-encoding gene (*P25*) of *Bombyx mori* was expressed in the anterior salivary gland after flies received the moth silk gene by *P* element-mediated transformation (Bello and Couble 1990; see Chapter 10 for a discussion of *P* element-mediated transformation). The *P25 B. mori* gene was appropriately expressed in the fly larval salivary glands, indicating that *Drosophila* salivary glands can

recognize *Bombyx* silk protein coding sequences and control their expression, despite the evolutionary divergence of flies and moths over 250 million years ago.

4.9.6. Transfer RNA Genes

More than 90 tRNAs have been identified during *Drosophila* development that are encoded by at least 670 genes, which can be divided into 60 separate groups. One to 18 tRNA genes are contained in each of 30 chromosomal sites, but there are no tandem repeats.

4.9.7. Vitellogenin Genes

Yolk proteins provide embryos with nutrients essential for growth within the egg. Most are phosphoglycoproteins and provide a source of amino acids, phosphate, lipids, and carbohydrates. The major yolk proteins are derived from **vitellogenins**, which are produced by the fat body and secreted for uptake by maturing oocytes. Vitellogenin gene structure and regulation has been studied in *Locusta migratoria*, the tobacco hornworm *Manduca sexta*, *Bombyx mori*, the boll weevil *Anthonomus grandis*, the Mediterranean fruit fly *Ceratitis capitata*, and *Drosophila* (Bownes 1986, Rina and Savakis 1991, Trewitt et al. 1992). The fat body of the mother is the primary producer of yolk proteins, but part are synthesized by the follicular epithelium of the ovary in *D. melanogaster*.

Yolk proteins in *Drosophila* consist of three polypeptides: YP1, YP2, and YP3. YP1 is expressed by the fat, body, and after posttranslational processing and glycosylation, the proteins are secreted into the hemolymph and delivered to the oocyte. YP2 is expressed in ovaries. The production and delivery of the three proteins are coordinately regulated and under the control of two hormones, 20-hydroxyecdysone and juvenile hormone (Bownes 1986). These two hormones regulate molting and metamorphosis during development as well.

Production of yolk proteins begins during the first day of *Drosophila* adult life. The production rate is high, with yolk proteins representing about one-third of the total proteins in the hemolymph. YP1 and YP2 are closely linked genes on the X chromosome, while YP3 also is sex-linked but more distant. YP1 and YP2 show much sequence homology and probably resulted from a fairly recent gene duplication event. Only one small intron is found in YP1 and YP2, and two in YP3. Extensive yolk protein synthesis in *Drosophila* is achieved because tissues are polytene and polyploid.

4.9.8. Transposable Elements

Transposable elements (TEs) are DNA sequences that can move (transpose) to new sites, invert, and undergo deletion or amplification. Transposable elements have been divided into two classes according to their structure and mechanism of transposition (Table 4.1). Class I elements transpose by reverse transcription of an RNA intermediate. Class I elements include elements related to retroviruses that have long terminal repeats (LTRs). They also include elements that have no long terminal repeats (non-LTR retrotransposons).

Class II elements transpose directly from DNA to DNA. They include elements with short inverted terminal repeats and have a coding region for a transposase. They also include elements with long inverted repeats. Many TEs have been discovered in *D. melanogaster* (Bowen and McDonald 2001). A diversity of TEs are known from other insects, as well (Table 4.2). A new class of TEs, rolling-circle transposons, have been found in eukaryotes (Kapitanov and Jurka 2001). Although rolling-circle TEs have not been found in arthropods

Table 4.1. A Classification of Transposable Elements by Their Method of Transposition

Class I Transposable Elements Transpose by Means of RNA Intermediates

A. Viral superfamily (retrovirus-like retrotransposons)
Have long direct terminal repeats (LTRs), encode reverse transcriptase from open reading frames (ORFs) in DNA between LTRs, able to generate 4-bp to 6-bp target site duplications, have no 3′ terminal poly(A) tract, are dispersed in genome.

Examples: *Copia*-like elements in *Drosophila melanogaster*
Gypsy-like elements in *D. melanogaster*
Pao in *Bombyx mori*

B. Nonviral superfamily (nonviral retroposons)
Have no terminal repeats, have ORFs, do not encode enzymes responsible for their transposition, have 3′ terminal poly(A) tract, are dispersed in genome.

Examples: *F* family in *D. melanogaster*
R2 retroposons in many insects
HeT-A retroposons in telomeres of *D. melanogaster*

Class II Transposable Elements Transpose Directly from DNA to DNA

All have a transposase and terminal inverted repeats (IRs)

A. With short inverted repeats (SIRs)
Examples: *P* and *hobo* in *D. melanogaster*
mariner in many insect species

B. With long inverted repeats (LIRs)
Example: *FB (foldback)* in *D. melanogaster*

(From Finnegan 1990, Robertson 1993, Xiong et al. 1993).

Table 4.2. Examples of the Diversity of Transposable Elements Identified from Arthropods Other Than *Drosophila*

Arthropod species	Element	Type	Reference(s)
Aedes aegypti	*Pony*	MITES, miniature inverted repeat TEs (Class II)	Tu 2000
	Lian	Non-LTR retrotransposon	Tu et al. 1998
Anopheles gambiae	T1	Non-LTR retrotransposon	Besansky 1990a,b
	Q	Non-LTR retrotransposon	Besansky et al. 1994
	Moose	LTR-retrotransposon	Biessmann et al. 1999
	Crusoe	*Tc1*-like (Class II)	Hill et al. 2001
	Vash, Guildenstern JuanAg	Non-LTR retrotransposons	Hill et al. 2001
	Ozymandias	LTR-retrotransposon	Hill et al. 2001
	8 novel families	MITES (Class II)	Tu 2000
Anopheles species	*Gypsy* family	LTR-retrotransposon	Cook et al. 2000
	Copia family	LTR-retrotransposon	
	LINE family	Non-LTR retrotransposon	
	Pao family	LTR-retrotransposon	
Apis mellifera	*G*	LINE-like retroposon	Kimura et al. 1993
	jockey	LINE-like retroposon	
Bactrocera tryoni	*Homer*	*hAT* TEs (Class II)	Pinkerton et al. 1999
	Homer-like *hopper*		Handler and Gomez 1997

continues

continued

Arthropod species	Element	Type	Reference(s)
Bombyx mori	*BmTc1*	*Tc1*-like element similar to that from *Caenorhabditis elegans* (Class II)	Mikitani et al. 2000
	jockey	LINE-like retroposon	Kimura et al. 1993
	Pao	Retrotransposon	Xiong et al. 1993, Abe et al. 2001
	L1Bm	Non-LTR retrotransposon	Ichimura et al. 1997; Abe et al. 1998
	TRAS1 and SART1	Telomeric repeat-associated retrotransposons	Okazaki et al. 1995, Takahashi and Fujiwara 1999
	Kabuki, Yokozuna	LTR-retrotransposons	Abe et al. 2000
	BMC1	Non-LTR retrotransposon	Abe et al. 2000
	Bm1	Retroposon	Abe et al. 2000
Chironomus thummi	TFB1	Foldback TE (Class II)	Hankeln and Schmidt 1990
Chironomus tentans	*NLRCt2*	Non-LTR retrotransposon	Blinov et al. 1997
Culex and *Aedes*	*Juan*	LINE-like retroposon	Mouches et al. 1991, 1992 Agarwal et al. 1993
C. pipiens	CM-gag	Similar to Het-A from *D. melanogaster*	Bensaadi-Merchermek et al. 1997
Heliothis virescens	*Hobo*-like	HAt elements (Class II)	DeVault and Narang 1994
Helicoverpa zea			
Lucilia cuprina	Lu-P1 and Lu-P2	Homologous to *P* elements (Class II)	Perkins and Howells 1992
	hermit	*hAT* transposable elements (Class II)	Coates et al. 1996
Lymantria dispar	LDT1	Non-LTR retrotransposon	Garner and Slavicek 1999
	Lydia	LTR-retrotransposon	Pfeifer et al. 2000
Many arthropods	*mariner*	Class II terminal repeats with a DNA intermediate	Robertson 1993 Robertson et al. 1992 Jeyaprakash and Hoy 1995
	R1 and R2 in rRNA genes	Non-LTR retroposons	Jakubczak et al. 1991 Bigot et al. 1992 Burke et al. 1993 Luan et al. 1993
Megaselia scalaris	TROMB	LTR-retrotransposon	Suck and Traut 2000
Musca domestica	*Hermes*	*hAT* TEs (Class II)	Warren et al. 1994 O'Brochta et al. 1996
Phlebotomus species	—	Non-LTR retrotransposon	Booth et al. 1994, 1996
Tribolium castaneum	*Woot*	LTR-retrotransposon	Beeman et al. 1996

to date, their distribution in plants and nematodes suggests they could have a broad host range.

At least half of all spontaneous mutations in *D. melanogaster* are due to insertions of TEs. For example, *P* elements in *D. melanogaster* cause excisions, chromosome rearrangements, and insertions. The foldback (FB) transposon is associated with deletions, inversions, reciprocal translocations, and insertional translocations in which normally

unique *Drosophila* DNA is flanked by two FB elements. All well-characterized, *highly unstable* genes in *D. melanogaster* were found unstable because they contained either the *P* element or FB elements (Berg and Howe 1989). Different TEs are found in *D. melanogaster* with different characteristics. For example, members of the HeT-A and TART families of TEs are found at telomeres and in centromeric heterochromatin and never in the euchromatin regions of chromosomes in *D. melanogaster* (Mason et al. 2000).

TEs could carry genetic information, regulate genes, or initiate genetic changes (Britten 1997, Miller et al. 1997, Shapiro 1999). Wilson (1993) suggested that TEs could lead to resistance to pesticides, although he did not provide any direct evidence for this. Agarwal et al. (1993) found a TE named Juan associated with amplification of the esterase gene in pesticide-resistant *Culex* mosquitoes, but a direct involvement in inducing gene amplification was not demonstrated. Waters et al. (1992) suggested the TE called *17.6* is involved in susceptibility to pesticides in *Drosophila* associated with a P450 gene. However, Delpuech et al. (1993) screened colonies of *D. melanogaster* and *D. simulans* from around the world and found no relationship between the presence or absence of *17.6* and resistance. One example of TEs containing genetic information may be found in *Drosophila hydei*. TEs and repetitive DNA sequences comprise the majority of the Y chromosome of *D. hydei*. Apparently the lampbrush-loop-forming fertility genes on the Y chromosome consist, at least in part, of retrotransposons of the *micropia* family (Huijser et al. 1988).

R1 and R2 are Class I TEs that lack long terminal repeats and were originally found in some of the 28S rRNA genes of the silk moth *Bombyx mori* and several Diptera. A survey suggests that R1 and R2 elements occur within the rRNA genes of many insects (Jakubczak et al. 1991). Forty-three of 47 species surveyed, including Odonata, Orthoptera, Dermaptera, Hemiptera, Homoptera, Coleoptera, Hymenoptera, Lepidoptera, and Diptera, contained the insertions in 5 to 50% of their 28S genes. The broad distribution of these elements raises the question of whether they could have been present in insects before their radiation more than 300 million years ago.

Very little is known about the origin and evolutionary history of TEs. A TE family might originate in a species, or TEs might be acquired by **horizontal or lateral transmission** from another species. Normally, DNA or RNA sequences are transmitted *vertically* from parent to progeny, but in horizontal transfer, DNA sequences are transferred laterally across species, taxonomic borders that were once thought to be inviolable (Daniels et al. 1990, Kidwell 1992, Plasterk et al. 1999). For example, the *hobo* element of *D. melanogaster* has a similar sequence to TEs from plants (*Activator* from corn and *Tam3* from snapdragon) (Calvi et al. 1991). Another element, *jockey*, identified from *D. melanogaster* has been found in the distantly related *D. funebris*, but not in species closely related to *funebris*. This again suggests that *jockey* moved horizontally from *D. melanogaster* into the genome of *D. funebris* (Mizrokhi and Mazo 1990).

A possible superstar at horizontal transfer may be *mariner*. Originally *mariner* was found in *Drosophila mauritiana* and several other species of *Drosophila*, as well as the moth *Hyalophora cecropia*. Subsequently, Robertson (1993) found that several types (subfamilies) of *mariner* are widespread in insects. It is found in other organisms as well. For example, *mariner* was found in the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae) (Jeyaprakash and Hoy 1995). It is likely that *mariner* has moved horizontally among diverse insect and mite species, although the frequency of horizontal transfer is infrequent on a human time scale (Robertson and Lampe 1995). Many *mariner* elements have degenerated and become inactive in the genomes of their hosts.

Relatively little information is available as to how TEs invade populations and the mechanisms involved in the first step of the invasion (Biemont et al. 1999). One of the best known examples is the invasion of *D. melanogaster* by the *P* element (see Chapter 9 for a review). Another evaluation of TE invasion was carried out by Biemont et al. (1999) in natural populations of *D. simulans*. *D. simulans* populations around the world are in the process of being invaded by a variety of TEs with whimsical names, including *1731*, *297*, *bel*, *blood*, *coral*, *F*, *flea*, *gypsy*, *HMS beagle*, *mariner*, *nomade*, *prygun*, *stalker*, and *zam*. Populations differ in the number and type of TEs, with some TEs absent in most populations, except for one or two populations which have high copy numbers. It is as yet unclear whether the new genetic variability that results from mobilization of TEs is adaptive. More research is needed to understand the mechanisms underlying the relationship between dramatic differences in TE copy number in different species and among natural populations and their environmental conditions. Biemont et al. (1999) suggest that the "initially selfish genes will surely appear more and more as'symbionts' that have played a major role in evolution and that may still provide genomic flexibility and variability for population adaptation."

If horizontal transmission of TEs occurs with some regularity, the implications are dramatic for evolutionary theory. Horizontal transfer risks could also influence regulations regarding the risks associated with releases of transgenic arthropods into the environment (Brosius 1991, Plasterk et al. 1999, Hoy 2000). As discussed in Chapter 9, studies of the transfer of *P* elements by a mite vector provide an intriguing glimpse at one possible mechanism by which TEs are able to move between species (Houck et al. 1991). Other possible vectors of TEs are insect viruses. Insect viruses may carry DNA from their hosts; a proportion of foreign DNA within insect viruses consists of TEs from the viruses' insect hosts (Fraser 1985).

4.10. Highly Repetitive DNA

Highly repeated DNA sequences with a uniform nucleotide composition can, upon fractionalization of the genomic DNA and separation by density gradient centrifugation, form one or more bands that are clearly different from the main band of DNA and from the smear created by other fragments of a more heterogeneous composition. These sequences are called **satellite DNA**.

Satellite DNA is sometimes described as minisatellite or microsatellite DNA, depending on the length of the repeated sequences. **Microsatellites** consist of tandem repeats of between 1 and 6 bp, often in long arrays; like other classes of repetitive DNA, microsatellites have high mutation rates (Bachtrog et al. 1999). Satellite DNA can comprise a large fraction of the arthropod genome. For example, in *Tribolium madens*, two satellite DNAs have been characterized: one is a 225-bp long monomer comprising 30% of the genome, and the second is a monomer of 711 bp, constituting 4% of the genome (Durajlija Zinic et al. 2000).

The role of highly repetitive sequences in genome evolution is not well understood (Ohno and Yomo 1991, Pardue and Hennig 1990), but highly repetitive sequences are associated with heterochromatin in the centromeres and telomeres and could be important in chromosome pairing. Telomeric DNA fragments have been isolated from *D. melanogaster*, *Chironomus*, and *Bombyx mori*. The sequence $(TTAGG)_n$ is found at the extreme terminal region of all *B. mori* chromosomes, as well as associated with the ends of chromosomes in the Isoptera, Orthoptera, Hymenoptera, Trichoptera, Mecoptera, some Coleoptera, Hemiptera,

and Lepidoptera (Okazaki et al. 1993). The sequence $(TTAGG)_n$ appears to be found only in arthropod telomeres, although not all arthropods have this telomeric sequence. *D. melanogaster* and some Coleoptera have different telomeric structures. The telomeres of *Drosophila* have TEs called HeT-A and TART in the subtelomeric region (Biessmann et al. 1993, Mason et al. 2000).

4.11. Producing Large Amounts of Protein in a Short Time: Gene Amplification and Gene Duplication

When it is desirable to produce large amounts of gene product in a short period of time, several mechanisms could be employed, including duplication of chromatids resulting in polyteny, polyploidy, hypertranscription, gene amplification, and gene duplication. **Hypertranscription** involves producing large amounts of gene product from a single copy of a chromosome and is the mechanism by which *D. melanogaster* males (which have only one X chromosome) produce as much gene product as females with two Xs (see Chapter 10 for more details). Definitions of gene amplification and gene duplication can be confusing (Edgar and Orr-Weaver 2001).

Gene amplification usually is defined as the replication of a gene (at a single locus) so that multiple copies can be transcribed at once. One way to visualize gene amplification is to imagine that gene amplification occurs by an "onion-skin model," in which a *segment* of the chromosome is replicated and multiple copies of that segment are transcribed while the rest of the chromosome retains its normal structure. Gene amplification originally was coined to describe the production in mammalian cell cultures of multiple copies of genes providing resistance to anticancer drugs. Gene amplification in cell cultures is associated with an initial low drug concentration, and the surviving cells are subjected to multiple rounds of selection with increasing concentrations of toxin. Gene amplification results in the production of more protein.

Gene amplification may be important to economic entomologists because some aphids and mosquitoes that are resistant to insecticides have amplified esterase genes, as described below, although the definition of gene amplification in these examples differs from that described for mammalian cells.

Gene duplication involves copying a gene multiple times; the copies may be maintained on the same chromosome in tandem array or be transferred to other chromosomes over evolutionary time.

4.11.1. Chorion Genes in *Drosophila* and Moths

Both gene amplification and gene duplication occur in chorion genes in *Drosophila* and the moths *Bombyx mori* and *Antheraea polyphemus* (Kafatos 1981, Eickbush and Burke 1985, Kafatos et al. 1986, Orr-Weaver 1991, Carminati et al. 1992). Analyses of these chorion genes resulted in significant advances in knowledge of the mechanisms of gene regulation and development.

In both *Drosophila* and silk moths, the egg is produced in the ovary, which consists of follicles composed of three cell types: 1) the oocyte, 2) a small number of nutritive nurse cells connected to the oocyte, and 3) follicular epithelial cells that surround the oocyte and nurse cells. There are approximately 1000 follicular epithelial cells per follicle in *Drosophila* and up to 10,000/oocyte in silk moths. These cells synthesize a complex mixture of proteins

and secrete them onto the surface of the oocyte to form the outer covering, or **chorion**. The chorion protects the embryo after fertilization and oviposition, preventing desiccation, yet enabling respiration to occur.

Drosophila and silk moth chorions are quite different. The *Drosophila* chorion is comparatively simple, with an endochorion and exochorion composed of six major and 14 minor proteins that are produced over approximately 5 hours. In silk moths, the number of genes and the time devoted to producing the chorion is much greater. There are three gene families in the wild silk moth *A. polyphemus*, and the same three families, plus two others, in the domesticated silk moth *B. mori*. Approximately 100 chorion proteins are produced during a period of approximately 51 hours in silk moths (Kafatos 1981).

Moth and fly chorion genes are organized differently. In *Drosophila*, large amounts of the chorion proteins are produced in a relatively short time (Lu et al. 2001). This is facilitated by *amplification* of the chorion genes. In *Drosophila*, there are two chorion gene clusters, 5 to 10 kb in size, each encoding tandemly oriented chorion genes. One gene cluster is found on the X and one on the third chromosome. Because each chorion gene cluster is represented only once in the haploid genome, the chorion proteins could not be synthesized quickly and in sufficiently large quantities unless gene amplification occurs. A 20-fold amplification of the chorion genes on the X chromosome, and an 80-fold amplification of the genes on chromosome 3 is found in follicle cells. Amplification is achieved by replicating the DNA segments at multiple replication origins (Heck and Spradling 1990). DNA amplification extends bidirectionally for a distance of up to 40–50 kb to produce a multiforked "onion-skin" structure that contains multiple copies of DNA containing the chorion genes (Figure 4.1).

Gene amplification also occurs in the chorion genes of the Mediterranean fruit fly *Ceratitis capitata*. The overall organization of the cluster is similar to that of *Drosophila*, with the same four genes maintained in tandem, in the same order, and with similar spacing

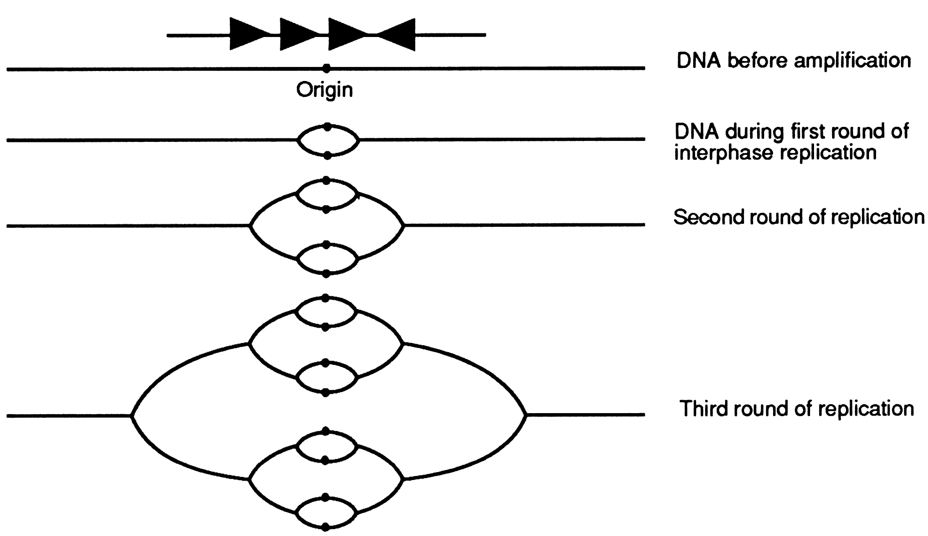


Figure 4.1. Amplification of the *Drosophila* chorion genes in follicle cells. The first three rounds of DNA replication at the 66D locus on chromosome 3. The three small arrows represent three well-characterized chorion genes in this cluster. The polarity of a fourth chorion gene and the precise location of the origin are unknown. The boundaries of the amplified DNA are much larger than the chorion protein transcription units within it.

(Konsolaki et al. 1990). Despite the divergence of *Drosophila* and *Ceratitis* family lineages approximately 120 million years ago, there is a high conservation in coding sequences and regulatory properties of their chorion genes.

Silk moth chorion proteins are produced over a longer time interval and involve larger numbers of genes that have probably arisen by **gene duplication**. More than 100 structural proteins are in the chorion of the silk moth *A. polyphemus*, which has an elaborate organization. In *B. mori*, the chorion genes are on chromosome 2 and consist of two segments that total more than 1000 kb of DNA. Subsets of the genes are expressed at different periods of choriogenesis (early, middle, late, very late), with the early proteins associated with framework formation, middle proteins with framework expansion, late proteins with densification, and very late proteins with surface sculpturing of the chorion.

Silk moths have solved the problem of producing large amounts of protein quickly by gene duplication. Silk moth chorion genes are found in multiple copies of divergently transcribed, coordinately expressed pairs (Kafatos et al. 1986). For example, all members from each of two late gene families are arranged in 15 pairs on a 140-kb segment. The members of each family have a high degree of sequence homology, although they are not identical. This homology could be maintained by a process called **concerted evolution** (Eickbush and Burke 1985). Concerted evolution often occurs in multigene families and could be maintained by two mechanisms: either unequal crossing over or gene conversion. Unequal crossing over may occur between the two sister chromatids of a chromosome during mitosis of a germ-line cell or between two homologous chromosomes at meiosis. It is a reciprocal recombination that results in a sequence duplication in one chromatid or chromosome and a corresponding deletion in the other. As a result of unequal exchange, daughter chromosomes become more homogeneous than the parental chromosomes. If the process is repeated, the numbers of each variant repeat on a chromosome will fluctuate with time, and eventually one will become dominant in the family. **Gene conversion** is a nonreciprocal recombination process in which two sequences interact so that one is converted by the other.

Despite the very different organization of chorion genes in *Drosophila* and silk moths, silk moth chorion genes can function in *D. melanogaster*. Moth chorion genes were cloned into *P*-element vectors and inserted into the *D. melanogaster* germ line (Mitsialis and Kafatos 1985). (Chapter 8 describes the methods employed in inserting moth genes into *Drosophila* using a *P* element that has been modified to carry exogenous genes.) Analysis of RNAs from transformed flies indicated that moth genes are expressed in an appropriate manner in the correct sex, tissue, and time in *D. melanogaster*. Fly and moth lineages diverged over 250 million years ago, yet regulatory elements conferring sex, tissue, and temporal specificity of gene expression must have been conserved. Chorion gene promoter sequences from the silk moths *Antheraea pernyi* and *A. polyphemus* also functioned in *D. melanogaster* after *P* element-mediated transformation, although some regulatory interactions had diversified (Mitsialis et al. 1989).

4.11.2. Insecticide Resistance

With the availability of molecular genetic techniques, geneticists have identified a new mechanism by which insects become resistant to pesticides (Mouches et al. 1990, Devonshire and Field 1991, Pasteur and Raymond 1996, Field 2000, Paton et al. 2000). Amplification of esterase genes in the aphid *Myzus persicae* and the mosquito *Culex pipiens quinquefasciatus* results in many identical gene copies present in tandem arrays in each cell.

Whether exposure to pesticides can induce resistance in insects by gene amplification is an interesting question. It has long been assumed that pesticide resistance in insects is due to the presence of rare alleles in populations that are selected for by pesticide applications (preadaptive mutations). However, amplification of genes in mammalian cells, plants, yeast, and microorganisms has been shown to occur in response to exposure to toxins. For example, amplification of the dihydrofolate reductase gene in mammalian cells in tissue culture occurs in response to exposure to methotrexate. A 100-fold amplification in a cholinesterase gene in two generations of a human family subjected to prolonged exposure to parathion has been demonstrated and could be due to genetic changes induced by prolonged exposure to this insecticide (Prody et al. 1989). Cultures of mosquito cells selected with methotrexate also became amplified (Fallon 1984). Thus, insecticide resistances due to gene amplification in insects could, at least in some cases, be induced by exposure to insecticides.

There are several mechanisms by which amplified genes are generated. These include gene duplication by random unequal crossing over between sister chromatids, with a subsequent misalignment resulting in increased numbers of tandem repeats with intervening spacer DNA (Stark and Wahl 1984). Another model involves replication of DNA more than once at the same origin of replication within a cell cycle (repeated replication model), which generates multiple unattached DNA molecules that either are released or are integrated into the chromosome by end-to-end ligation and recombination. The repeated replication model is consistent with the sudden appearance of many gene copies and the initiation of chromosome breaks or translocations, which are often found associated with gene amplification. In both aphids and mosquitoes the increase in gene copy number appeared to occur in a stepwise manner.

In both aphids and mosquitoes, traditional analyses of the mode of inheritance of resistance indicate that high esterase activity is inherited as a single factor, which is expected if the amplified genes are located on the same chromosome and inherited as a unit. However, in both aphids and mosquitoes, resistance can be unstable in the absence of selection. In the parthenogenic aphid clones, resistance is usually stable for long periods but can be lost in some progeny of some clones. This reversion is associated with clones that carry the translocation, and the revertant lines can often be reselected for resistance. Because reversion involves loss of elevated esterase production and mRNA, but not of the amplified genes, it is likely that control over gene transcription is responsible for the loss of resistance.

4.12. Multiple Genomes in Insects: What Is the "Biological Individual?"

Eukaryotes typically have a nuclear genome and a mitochondrial genome. Mitochondria are now generally accepted to be microbial **symbionts** that were modified after a long process of evolution within eukaryotic cells (Gray 1989, Martin 1999). Mitochondria retain a distinctive genome that is replicated and expressed, but mitochondria are incapable of independent existence. In the course of evolution some mitochondrial genes were transferred to the nuclear genome of its host.

4.12.1. Multiple Symbionts

In addition to mitochondria, insects have intimate intra- and extracellular relationships with a diverse array of organisms including viruses, bacteria, yeasts, and rickettsia (Schwemmler

and Gassner 1989, Douglas 1992). The details of the relationship between the host and these microorganisms usually are unknown, but we are learning more with the aid of molecular tools.

For example, the rice weevil *Sitophilus oryzae* (Rhynchophoridae) has four intracellular genomes that are involved in the weevil's biology. These are nuclear, mitochondrial, principal endosymbiont, and *Wolbachia* (Heddi et al. 1999, 2001). The principal endosymbiont is found (3×10^3 bacteria/cell) in specialized bacteriocytes. A total of 3×10^6 bacteria are found in each weevil, which is 10-fold more cells than there are beetle cells (Heddi et al. 2001). These symbionts induce the specific differentiation of the bacteriocytes and increase mitochondrial oxidative phosphorylation through the supply of pantothenic acid and riboflavin. Their elimination impairs many important physiological traits, including flight ability. This weevil supports the "serial endosymbiotic theory"; according to this view, endosymbiosis did not occur just once in eukaryotic evolution with the origin of a nucleus, or even twice, when an anaerobic protist acquired a respiring bacterium to give rise to the mitochondrion. The acquisition of genomes by eukaryotic cells "continues today in the multicellular organism" (Heddi et al. 1999). The rice weevil gut symbiont allowed the weevil to colonize cereal plants because it supplied vitamins. Heddi et al. (1999) "consider symbiosis in the rice weevil a sophisticated mechanism for acquiring new sets of genes."

Symbionts may possess metabolic capabilities that the insect host lacks, and the insect uses these capabilities to survive on poor or unbalanced diets (Douglas 1998). Under such circumstances, the insect and microbe relationship is often required. Many insects freed of their symbionts grow slowly and produce few or no progeny; many microorganisms cannot grow outside their insect host, indicating the relationship is a long and intimate one. Intracellular symbionts are found in the Anoplura, Mallophaga, Isoptera, Orthoptera, Homoptera, Coleoptera, Diptera, and Hymenoptera. The amazing diversity of relationships and organisms involved in these relationships with insects has raised many questions, but provided few clear-cut answers, in large part because most symbionts cannot be cultured outside their hosts. Many microorganisms are contained in special structures and transmitted by a highly specific method, including transovarial transmission, to progeny. Transmission also can occur when larvae feed on contaminated egg shells or feces.

Some insect species contain several different types of symbionts in different tissues, including the gut, Malpighian tubules, fat body, or gonads. Bacteroids, spiroplasmas, rickettsia, mycoplasmas, or virus-like symbionts are found in dipteran testes, ovaries, pole cells, nurse cells, and gut wall cells. Endocytobiosis in scale insects (Homoptera) is particularly diverse with almost 20 different types of associations described so far. In the leafhopper *Euscelidium variegatus*, specific bacteria are thought to be essential for normal growth and development, breaking down uric acid in the host cells and synthesizing amino acids and vitamins. Symbionts are involved in normal egg development of *E. variegatus*; embryos artificially lacking symbionts fail to develop normal abdomens. It is hypothesized that some genes from this microorganism have been transferred to the nuclear genome of *E. variegatus* in a manner parallel to that of mitochondria.

Some insects lacking their symbionts are apparently completely normal. For example, in the beetle family Cerambycidae, all of which live in wood, some species have symbionts while others lack them. The hypothesis that symbionts supply a nutrient deficiency in the insect's diet thus appears to be simplistic; some insects feeding on a well-balanced diet have symbionts.

In some cases, rickettsia-like symbionts increase the likelihood that an insect vector can transmit (vector) a disease. For example, rickettsia-like organisms in the tsetse fly *Glossina*

morsitans morsitans affect infection by the sleeping sickness trypanosomes (Welburn et al. 1993). The rickettsia-like organisms produce endochitinases in the tsetse gut that inhibit lectins in newly emerged adults. Tsetse flies lacking the rickettsia-like organisms are less susceptible to trypanosomes (are refractory), and transmit the disease less often.

Microbial symbionts are common in insects, but a full understanding of their genetic and evolutionary role remains to be determined. In the few cases that have been well studied, a genetic interplay between insect host and symbiont occurs, each supplies factors to the other, and the microorganism has specific means of movement and relocation within the insect. A symbiont must be recognized by the insect as "self" rather than as foreign or they would be subject to the insect's immune system. Our understanding of how microorganisms have become incorporated into insect organ tissues and cells remains fragmentary, but is advancing with the use of molecular tools (Schwemmler and Gassner 1989, Moran and Baumann 2000).

Perhaps the most unusual recent discovery is that there are "bugs within bugs within mealybugs" (von Dohlen et al. 2001). Mealybugs (Pseudococcidae) have endosymbionts that live within the cytoplasm of large, polyploid host cells within a specialized structure (bacteriome). These symbionts provide nutrients to their hosts. The relationship between homopteran insects and these primary endosymbionts is ancient, perhaps dating to the origins of the families or superfamilies 100 to 250 million years ago (von Dohlen et al. 2001). The mealybug hosts, *Planococcus citri*, package their intracellular endosymbionts into mucus-filled spheres which surround the host cell nucleus and occupy most of the cytoplasm. These spheres are structurally unlike eukaryotic cell vesicles. von Dohlen et al. (2001) were able to demonstrate that the mealybug host cells actually harbor two types of Proteobacteria. The two bacteria are not co-inhabitants of the spheres. Rather, the spheres themselves are β-proteobacteria and the γ-proteobacteria are living inside them. This was the first report of an intracellular symbiosis in which one bacterium lives within another. von Dohlen et al. (2001) hypothesized that the internalization of the one bacterium by the second may facilitate the exchange of genes and gene products which could slow or reverse the genetic degradation that is common to organelles or long-term intracellular symbionts over evolutionary time.

4.12.2. *Wolbachia*

A genus of α-proteobacteria called *Wolbachia* is commonly found in arthropods (Werren et al. 1995, Rigaud and Rousset 1996, O'Neill et al. 1997, Jeyaprakash and Hoy 2000). *Wolbachia* are intracellular gram-negative rods that cannot be cultured easily outside their hosts. *Wolbachia* infection rates range from approximately 17 to as much as 76% of all arthropod species (Werren et al. 1995, Jeyaprakash and Hoy 2000). *Wolbachia* also have been found in crustaceans (Rigaud 1999) and nematodes (Bandi et al. 1999, Bazzocchi et al. 2000). An understanding of their physiological and phenotypic effects on their hosts is still being developed. The phylogeny of *Wolbachia* in nematodes is congruent with the phylogeny of their hosts, suggesting they share a long coevolutionary history. That is not true for arthropods, where it appears a great deal of horizontal transfer has occurred. Treatment with the antibiotic tetracycline inhibits normal reproduction and development of filarial nematodes that harbor *Wolbachia*, suggesting that *Wolbachia* are necessary to the nematode (Langworthy et al. 2000).

Wolbachia have been implicated as the cause of both alterations in sex ratio (resulting in thelytoky and male killing), which will be discussed further in the chapter on

sex determination (Chapter 10), and **cytoplasmic incompatibility** in arthropods. Some *Wolbachia* improve fertility or vigor, while others appear to decrease these traits in their hosts.

The molecular mechanism(s) by which reproductive incompatibility is induced by *Wolbachia* are hypothesized to be due to *Wolbachia*'s ability to modify sperm. This hypothesis suggests that paternal chromosomes are modified during spermatogenesis by *Wolbachia* and this modification is "rescued" in eggs of females infected with the same strain of *Wolbachia* during fertilization. If, however, the female is not infected with *Wolbachia* and mates with an infected male or male infected with a different strain of *Wolbachia*, then the embryos die (Figure. 4.2A). Some *Wolbachia* strains have been identified that fail to modify sperm but can rescue the modification in eggs of other *Wolbachia* strains (Bourtzis et al. 1998).

Cytoplasmic incompatibility caused by *Wolbachia* may be partial or complete. Sometimes incompatibility is found in both reciprocal crosses ($A \times B$ and $B \times A$, **bidirectional incompatibility**), perhaps due to the presence of different strains of *Wolbachia* in each population. Incompatibility is more often found in one reciprocal cross ($A \times B$ or $B \times A$, **unidirectional incompatibility**). Cytoplasmic incompatibility typically is incomplete (less than 100%), perhaps because of inefficient transfer of *Wolbachia* to all progeny or to differences in the titer of *Wolbachia*. Such differences in titer could occur naturally if the infected insects encounter antibiotics in their environment or if they experience high temperatures (typically $>30°C$) (Snook et al. 2000).

Some insects appear to have *Wolbachia* only in their germ-line tissues (ovaries and testes) while others have *Wolbachia* in somatic tissues (Dobson et al. 1999). Large numbers of *Wolbachia* have been found in ovaries and testes of populations with cytoplasmic incompatibilities. Incompatible strains have been converted to compatible by treating the colonies with heat or antibiotics, which eliminates or greatly reduces the *Wolbachia* population.

Wolbachia can be transferred to new populations experimentally by microinjecting infected egg cytoplasm into uninfected eggs. Transinfected strains of *D. simulans* and *D. melanogaster* with high titers of *Wolbachia* exhibited cytoplasmic incompatibilities at high levels, but those with low titers exhibited low levels of incompatibility, suggesting that a threshold level of infection is required and that host factors may determine the density of the *Wolbachia* in the host (Boyle et al. 1993).

Wolbachia have been identified in at least 70 species of parasitic Hymenoptera, including species in the Aphelinidae, Encyrtidae, Eulophidae, Pteromalidae, Torymidae, Trichogrammatidae, Cynipidae, Eucoilidae, Braconidae, Ichneumonidae and Proctotrupoidae, and in three dipteran parasitoids (Tachinidae) (Cook and Butcher 1999). Both cytoplasmic incompatibility and induction of parthenogenesis in these parasitoids may be caused by *Wolbachia*. Many hymenopteran parasitoids have both bisexual (arrhenotokous) and unisexual strains consisting only of females (thelytoky), probably due to the presence of *Wolbachia*.

Phylogenetic analysis suggests that the *Wolbachia* common ancestor evolved between 80 and 100 million years ago (O'Neill et al. 1992), whereas the arthropod common ancestor occurred at least 200 million years earlier. Thus, *Wolbachia* probably have invaded arthropods through **horizontal transmission** (Heath et al. 1999, O'Neill et al. 1992, Jeyaprakash and Hoy 2000). Some arthropods have been found to have double or even triple infections of *Wolbachia*. The effects of these multiple infections usually are unknown.

Several methods have been proposed as mechanisms for horizontal transfer, including the movement of *Wolbachia* from host arthropods to their parasitoids. Heath et al. (1999)

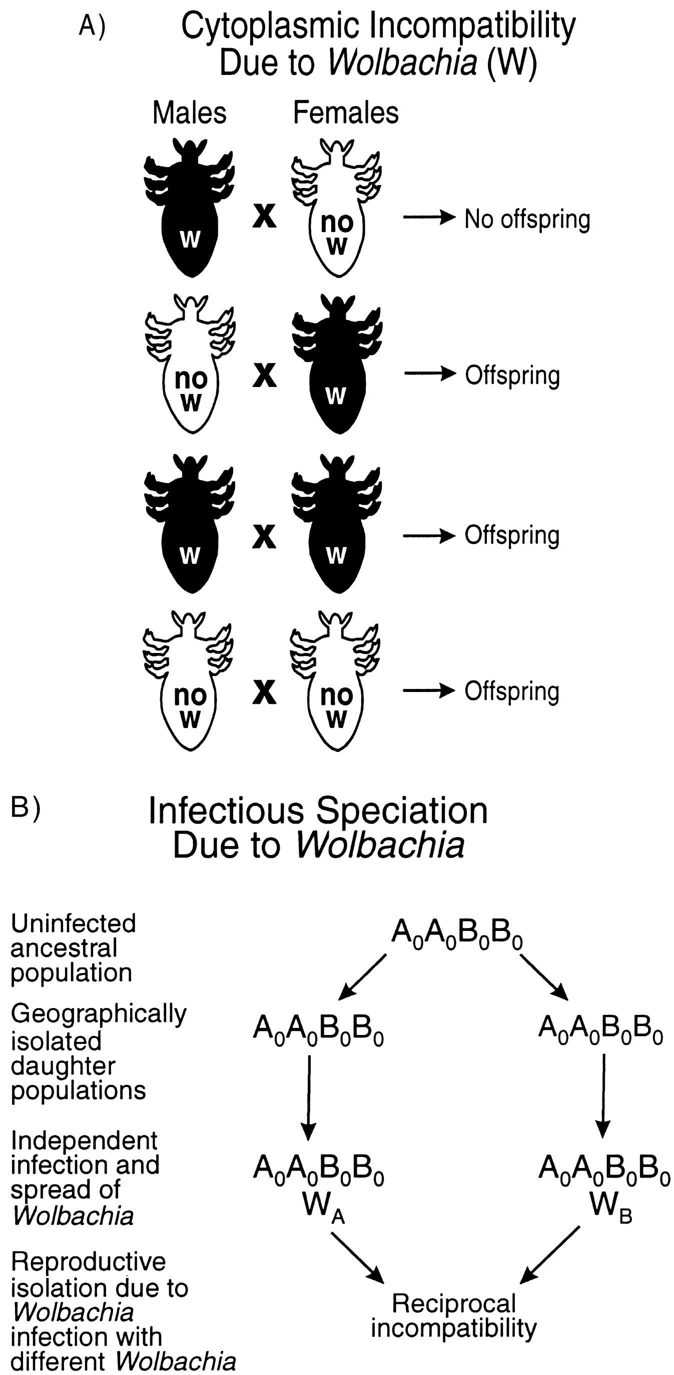


Figure 4.2. A) Cytoplasmic incompatibility due to *Wolbachia* between different individuals or populations can result in a failure to produce progeny when a *Wolbachia*-infected male mates with an uninfected female. B) If isolated populations of a species become infected with different types of *Wolbachia* (W_A and W_B), then these populations could become reproductively isolated if they later come into contact.

experimentally transferred *Wolbachia* from *Drosophila simulans* to a novel host, its endoparasitoid *Leptopilina boulardi*, 0.711% of the time. The *Wolbachia* infection rate diminished during subsequent vertical transmission to the F_2 and F_3 generations, perhaps because of poor maternal transmission (unstable vertical transmission).

Experimental microinjection (artificial horizontal transfer) of *Wolbachia* from the parasitoid *Muscidifurax uniraptor* into its host *D. simulans* resulted in a temporary infection, but no specific phenotypic effects were observed (van Meer and Stouthamer 1999). These results suggest that host–symbiont interactions are important for successful establishment of a *Wolbachia* infection in a new host. It is clear that *Wolbachia* has successfully bridged large phylogenetic distances in its horizontal movements over evolutionary time.

The availability of PCR primers for *Wolbachia* genes revolutionized the study of the distribution and evolution of *Wolbachia*. The *Wolbachia* genome project (see below) will further revolutionize such studies. Based on a phylogeny developed using the *ftsZ* gene, *Wolbachia* infecting arthropods have been divided into Groups A and B, which are estimated to have diverged from each other 58 to 67 million years ago (Werren et al. 1995). Phylogenies based on *wsp* gene sequences have yielded more groups, indicating that considerable genetic variation exists (Zhou et al. 1998, van Meer and Stouthamer 1999, Jeyaprakash and Hoy 2000). It is unclear whether these *Wolbachia* groups are strains or species.

Wolbachia may have a role in speciation of arthropods by generating reproductive isolation (Rokas 2000), although some argue that *Wolbachia*'s role(s) remain unproved (Hurst and Schilthuizen 1998). Typically, *Wolbachia* cause unidirectional cytoplasmic incompatibility when a *Wolbachia*-infected male mates with an uninfected female (Figure 4.2A). The eggs or embryos of such matings die, resulting in a fitness cost to uninfected females, which over time results in the infected cytotype becoming fixed in the population. A problem with this speciation hypothesis is that *Wolbachia* are not transmitted 100% of the time from a female to her progeny, so that some progeny will be produced that are compatible. Secondly, incompatibility is not completely expressed (incomplete penetrance of the trait) when infected males and uninfected females mate in natural populations (perhaps due to differences in the titer of the *Wolbachia* within individuals). Furthermore, selection on both the host and *Wolbachia* may favor reduced penetrance of the incompatibility phenotype or loss of *Wolbachia*. This could lead to a situation in which there is no gene flow to some gene flow (Hurst and Schilthuizen 1998). Thus, unidirectional incompatibility caused by *Wolbachia* may be insufficient to cause the reproductive barriers that could lead to speciation. Additional factors, such as **hybrid sterility** (sterility of the hybrid when crossed with either of the parental species) and **hybrid breakdown** (the inviability or sterility of progeny resulting from a backcross of hybrid progeny with either of the parental species), may be necessary (Shoemaker et al. 1999). *Wolbachia* may enhance the speciation rate by acting in conjunction with behavioral isolation.

A second speciation mechanism associated with *Wolbachia* may be by the induction of thelytoky (reproduction by females only), as has been found in a number of hymenopteran parasitoids such as *Encarsia formosa*. Populations of *Encarsia* no longer have males, so that populations essentially become clonal and over time could differentiate genetically (Cook and Butcher 1999).

A third potential *Wolbachia* speciation mechanism is by bidirectional incompatibility; if a population is infected with two different strains of *Wolbachia* that are incompatible with each other, then the incompatibility could act as a postzygotic reproductive barrier, as has been suggested for the species complex of *Nasonia* (Hymenoptera)(Figure. 4.2B). Thus, how *Wolbachia* are maintained in populations has considerable theoretical and practical

importance. *Wolbachia* have been proposed as vectors for transforming their hosts, as well as mechanisms for driving genes into populations in genetic manipulation projects for improved pest control (see Chapter 14 for additional discussion of this topic).

The interest in the biology and evolution of *Wolbachia*, with its fascinating effects on reproductive isolation (thus potentially having effects on speciation), sex ratio, feminization, and male killing, has led to the development of a *Wolbachia* genome project (Bandi et al. 1999). Four groups of *Wolbachia* are targets: one from the filarial worm *Brugia malayi* (group D); one from *D. simulans* associated with cytoplasmic incompatibility; one associated with parthenogenesis in the parasitoid *Muscidifurax uniraptor*, and one associated with feminization in the crustacean *Armadillidium vulgare*. Genome sizes for six different *Wolbachia* strains from these hosts were determined by pulsed-field gel electrophoresis (Sun et al. 2001). The *Wolbachia* genomes are circular and range in size from 0.95 to 1.66 Mb, which is considerably smaller than the genomes of free-living bacteria such as *Escherichia coli* (4.7 Mb).

Despite the wealth of information obtained about *Wolbachia* within the past few years, our understanding of the role of *Wolbachia* in arthropod biology and evolution probably remains fragmentary. For example, some *Wolbachia* in arthropods have been shown to contain bacteriophages named WO (Masui et al. 1999, 2000). A phylogenetic analysis of different WOs from several *Wolbachia* strains yielded a tree that was not congruent with the phylogeny of the *Wolbachia*, suggesting that the phages were active and horizontally transmitted among the various *Wolbachia*. Masui et al. (2000) speculated that, because all *Wolbachia* strains they examined contain WO, the phage might have been associated with *Wolbachia* for a very long time, conferring some benefit on its microbial hosts.

4.12.3. Polydnaviruses in Parasitoids

A particularly interesting example of an intimate relationship between insects and symbionts is illustrated by the relationship between polydnaviruses and parasitoids. The **polydnaviruses** are found only in the Braconidae and Ichneumonidae among the parasitic Hymenoptera (Krell 1991, Fleming 1992, Stoltz and Whitfield 1992). Polydnaviruses are symbiotic proviruses that have double-stranded circular DNA genomes; they are literally "poly DNA-viruses," having segmented genomes composed of several circular DNA molecules. For example, the viral genome within *Campoletis sonorensis* consists of 28 DNA molecules ranging in size from approximately 5.5 to 21 kb, with the total genome size equal to approximately 150 kb.

Polydnaviruses ensure that some species of braconids and ichneumonids (= parasitoids) are able to successfully parasitize their insect hosts. At least 50 species of parasitoids contain polydnaviruses (Stoltz and Whitfield 1992), and more than 30,000 species are thought to carry them (Shelby and Webb 1999). Genera of parasitoids containing polydnaviruses appear to be more speciose and have broader host ranges than sibling groups lacking them, suggesting the viruses contribute to the evolutionary success of their hosts (Shelby and Webb 1999). The two polydnaviral groups, Ichnoviridae and Brachoviridae, are phylogenetically and morphologically distinct and use different mechanisms to inhibit host immunity and development (Webb 1998). The association between braconid parasitoids and their viruses appears to have lasted at least 60 million years (Whitfield 1997).

Polydnaviruses replicate only in braconid or ichneumonid wasp ovaries and are secreted into the oviducts from where, during oviposition, they are injected into host lepidopteran larvae. The viruses appear to be vertically transmitted and integrated into the chromosome of

the wasp (Fleming and Summers 1991). Each wasp species appears to carry a polydnavirus characteristic of that species. If one species within a particular genus carries a polydnavirus, all others within the genus are likely to do so (Stoltz and Whitfield 1992).

Insects possess immune mechanisms that protect them from microorganisms, other invertebrates, and abiotic materials (Hultmark 1993, Gillespie et al. 1997). Protection occurs through constitutive factors or by inducible humoral and cellular responses. Many behavioral, morphological, nutritional, and endocrine factors determine whether the interactions between a host and a parasitoid will lead to development of the parasitoid or to its destruction (Fleming 1992). Polydnaviruses alter the host insect's neuroendocrine and immune responses, preventing encapsulation of parasitoid eggs and larvae by host hemocytes, and influence development of the host to benefit the parasitoid (Webb and Cui 1998, Shelby and Webb 1999). The virus replicates asymptomatically in the parasitoid but causes a pathogenic infection in the lepidopteran host (Webb and Cui 1998). The virus alone can induce altered immune responses in some hosts, but in other hosts the venom injected by the wasp also must be present for the full effect of the virus to occur. Parasitoid wasps thus appear to benefit significantly from the polydnaviruses that replicate in their reproductive tracts. The virus also clearly benefits if the parasitoid is able to reproduce, because polydnaviruses are known to replicate only within their host parasitoids.

The polydnavirus–parasitoid–lepidopteran host system provides an unusual example of an obligate mutualistic association between a virus and a parasitoid that functions to the detriment of the parasitoids' lepidopteran host. The origin of polydnaviruses is unknown, as is how they became established in the parasitoid genome. Beckage (1998) speculated that polydnaviruses may have potential value in agricultural pest management programs if genetically engineered pathogens (viruses, bacteria, fungi) containing polydnavirus genes could produce products that suppressed the target pest's immune system. Alternatively, genetically engineered parasitoids could be developed that exhibit a modified host range, making them more effective in controlling pests.

4.12.4. Gut Symbionts in Arthropods

Insects may contain complex and diverse societies of microbes in their guts, yet relatively little is known about how these resident microbes shape the physiology of their hosts (Cazemier et al. 1997, Kaufman et al. 2000). The primary habitat for microorganisms associated with insects is the hindgut. The termite gut is one of the better studied examples, and molecular tools are improving our ability to resolve the taxonomy of the complex relationships among termite gut symbionts.

The hindguts of termites can be compared to small bioreactors where wood and litter are degraded, with the help of symbiotic microorganisms, to provide nutrients. The hindgut of termites is a structured environment with distinct microhabitats (Brune and Friedrich 2000). The dense gut microbiota includes organisms from the Bacteria, Archaea, eukaryotes, and yeasts. These diverse organisms do not occur randomly within the gut but may be suspended in the gut contents, located within or on the surface of flagellates, or attached to the gut wall. The identity, exact number, and location of most is inadequately known because these organisms are difficult to culture. Molecular tools are providing significant new information. The spirochaetes, which account for as many as 50% of the organisms present in some termites, are a distinct phylum within the bacterial domain, but relatively little is known about them. One molecular analysis of spirochaetes in the termite *Reticulitermes flavipes* suggested there are at least 21 previously unknown species

of *Treponema* (Lilburn et al. 1999). The authors concluded that the long-recognized and striking morphological diversity of termite gut spirochaetes is paralleled by their genetic diversity, which could reflect substantial physiological diversity (Lilburn et al. 1999).

Omnivorous cockroaches also have gut microbial communities, but the associations are less interdependent than those of termites. As in termites, the gut microbial communities in cockroaches anaerobically degrade plant polymers and include hydrogen-consuming bacteria, especially methanogens. The densities of these microorganisms can be enormous; for example, 5×10^{12} bacteria per ml were found in the hindgut of the cockroach *Periplaneta americana* (Cazemier et al. 1997).

Antlions (Myrmeleontidae) suck out the body fluids of their prey after first paralyzing them with a toxin produced by salivary gland secretions produced by bacteria located in the salivary glands. The paralyzing toxin produced by these bacterial endosymbionts is a homolog of GroEL, a heat-shock protein that functions as a molecular chaperone in *E. coli* (Yoshida et al. 2001). In the antlion, the GroEL protein may act on receptors in prey insects to induce paralysis. The antlion symbionts perhaps evolved this nonchaperone function to establish a mutually beneficial antlion–bacterium relationship. Yoshida et al. (2001) speculated that insecticidal proteins may be produced by other endosymbionts to help additional fluid-feeding predatory insects.

Tsetse flies (Glossinidae) are vectors of African sleeping sickness disease in humans and animals. Microorganisms associated with these flies, which are blood feeders, are responsible for nutrients not found in their restricted diet. Different microorganisms have been found in the midgut, hemolymph, fat body, and ovaries. Until molecular techniques were used, their taxonomic status was unresolved (Aksoy 2000). Now we know that at least three different microorganisms are present: the primary (P) symbiont, *Wigglesworthia glossinidia*, is an intracellular symbiont residing in specialized epithelial cells that form a special U-shaped organ (bacteriome) in the anterior gut. The secondary gut symbiont, *Sodalis glossinidius*, is present in midgut cells. The third, *Wolbachia*, is found in reproductive tissues. Tsetse females are viviparous, retaining each egg within the uterus where it hatches. The larva matures there and is born as a fully developed third-instar larva. During its intrauterine life, the larva receives nutrients and both of the gut symbionts from its mother via milk-gland secretions; the *Wolbachia* are transmitted transovarially. Efforts to eliminate tsetse symbionts with antibiotics result in retarded growth and a decrease in egg production. Because it is difficult to eliminate only one symbiont at a time, it is difficult to decipher the role each plays. However, the gut symbionts supply B-complex vitamins, and *Sodalis* also produces a chitinase, which appears responsible for increasing the susceptibility of its host to the sleeping sickness trypanosome (Aksoy 2000). Analysis of the *Wigglesworthia* and *Sodalis* genomes indicates that they each form a distinct lineage in the Proteobacteria. Molecular analyses suggest that a tsetse ancestor was infected with a *Wigglesworthia* and from this ancestral pair evolved the tsetse species and *Wigglesworthia* strains existing today. No evidence was found for horizontal transfer of *Wigglesworthia* symbionts between tsetse species. *Sodalis* infections might represent recent independent acquisition by each tsetse species or multiple horizontal transfers between tsetse species.

Among the best-studied endosymbionts of insects is *Buchnera aphidicola*, a bacteriocyte-associated endosymbiont of aphids (Baumann et al. 1997, Douglas 1998, Moran and Baumann 2000). Its complete genome has been sequenced (Shigenobu et al. 2000). *Buchnera* is found in huge cells (bacteriocytes) in most of the 4400 aphid species, supplying the aphids with essential amino acids. In return, *Buchnera* is given a stable and

nutrient-rich environment. Aphids become sterile or die if their symbionts are eliminated. The aphid–*Buchnera* relationship has been stable for up to 250 million years, and about 9% of the *Buchnera* genome is devoted to producing essential amino acids for the aphid. Genes for nonessential amino acids are absent in *Buchnera*, and this symbiont depends on its aphid host for these, making *Buchnera* and the aphid codependent.

Analyses of different aphid species and their *Buchnera* symbionts indicate that vertical transmission of the symbionts has occurred from the time of the common ancestor of aphids, approximately 150 to 250 million years ago (Moran and Baumann 2000). Thus, there is "phylogenetic congruence with hosts, implying co-speciation," and there is no evidence of horizontal transfer, even within a single aphid species (Moran and Baumann 2000). In many *Buchnera* lineages, genes involved in tryptophan and leucine biosynthesis are present on plasmids rather than in the *Buchnera* genome. The location of these genes on plasmids allows increased gene expression and, thus, increased benefit to their aphid hosts. The number of copies of the plasmids appears to vary across *Buchnera* in different aphid lineages, perhaps reflecting coordinated, adaptive adjustment to the nutritional needs of the different aphid hosts. The genome of *Buchnera* is unusual when compared to the free-living bacterium *E. coli*. First, the sequences are very AT-biased (about 28% GC). Second, DNA sequences evolve faster in *Buchnera* than in free-living relatives. Third, the genome of *Buchnera* (from *A. pisum*) is reduced to about 650 kb, which is about one-seventh of the genome size of *E. coli*. *Buchnera* appears to contain only a subset of about 600 of the 4500 genes present in an *E. coli*-like ancestor.

Remarkably, it appears that each *Buchnera* contains 50 to 200 chromosomes, with the number of copies varying with the life-cycle stage of the host. Chromosome amplification may be used to vary the contribution of the symbiont to its host's nutrition (Komaki and Ishikawa 1999, 2000). The amplification of chromosome copy number to 200 copies/cell is very unusual in the microbial world; *E. coli* typically has one or two chromosomes per cell. The dramatic reduction in genome size of *Buchnera* and the extraordinary increase in genome copy number make this intracellular symbiont resemble eukaryotic cell organelles such as mitochondria and chloroplasts—which are evolutionary descendants of symbiotic bacteria (Komaki and Ishikawa 2000). *Buchnera* resemble these organelles also in that they are transmitted maternally between aphid generations.

A less intimate relationship between microbial genomes and insects is represented by the relationship between *Enterobacter agglomerans*, found in the gut of the apple maggot *Rhagoletis pomonella* (Lauzon et al. 2000). Enterobacteriaceae are the most common microorganisms associated with the apple maggot in the gut and female reproductive organs, and there are suggestions the flies use the bacteria for some vital function(s) (Lauzon et al. 1998). In addition to *E. agglomerans*, *Klebsiella oxytoca* is found in the gut of *R. pomonella*, and both are most abundant in the esophageal bulb, crop, and midgut. These bacteria are found on host plants and other substrates in the environment. It appears that the bacteria provide usable nitrogen for *R. pomonella* and other tephritids by degrading purines and purine derivatives, making them facultative symbionts. The relationship between the *Enterobacter* and *Klebsiella* species is probably complex. Figure 4.3 illustrates the biofilm of *E. agglomerans* and *Klebsiella* in an adult *R. pomonella* midgut. A **biofilm** is a complex, structured community of microbes attached to surfaces. Microbial biofilms function as a cooperative consortium in a complex and coordinated manner (Davey and O'Toole 2000). The role of this biofilm in *R. pomonella* is under study (Lauzon et al. 1998).

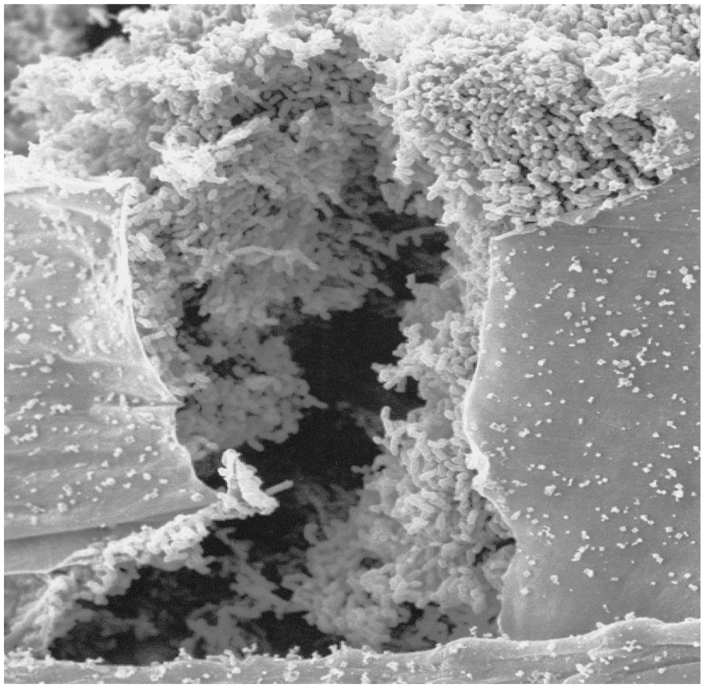


Figure 4.3. This scanning electron micrograph shows a biofilm of *Enterobacter agglomerans* and *Klebsiella* species in the midgut of the apple maggot *Rhagoletis pomonella*. (Photo kindly provided by C. R. Lauzon.)

4.13. Insect Development

Much of what we know about the genetics of development in insects has been learned by studying *Drosophila melanogaster* (Lawrence 1992, Wilkins 1993, Powell 1997, Otto 2000), although that is beginning to change (Klingler 1994). Extensive analyses of insect development have become feasible with the tools of molecular genetics, and thousands of papers have been published on the molecular genetics of development in *D. melanogaster*. Review articles and books have been published on this rapidly advancing field (Lawrence 1992, Wilkins 1993). A complete discussion of development is beyond the scope of this chapter. However, the following provides a brief outline of *D. melanogaster* embryonic development that will be useful in understanding sex determination, behavior, and *P* element-mediated transformation (Chapters 9, 10, and 11).

4.13.1. Oocyte Formation in *D. melanogaster*

A substantial amount of development of the insect embryo is determined in the oocyte, before oocyte (n) and sperm (n) pronuclei fuse to form an (2n) embryo. Oocyte formation in *D. melanogaster* is complex, involving both somatic and germ-line cells. The ovaries contain oocytes, which are formed from the pole cells, but the cells that surround each egg chamber and make up the walls of the egg chambers are derived from mesoderm (somatic tissues). The pro-oocyte arises in a set of cell divisions within the ovary from an oogonial stem cell. Each oogonial stem cell divides to give a daughter stem cell and

a cystoblast cell. The cystoblast cell gives rise to a set of 16 sister cells in four mitotic divisions, which provides a cyst. One of these 16 cells becomes the pro-oocyte, and eventually the oocyte, while its 15 sister cells become nurse cells whose function is to synthesize materials to supply the growing oocyte. The 16-cell cyst, surrounded by a layer of somatic cells, is termed the egg chamber. The final stages of egg chamber development involve covering the cyst with a monolayer of pre-follicle cells, which are somatic in origin. These 80 somatic cells divide an additional four times to give 1200 follicle cells which cover each cyst.

Initially *Drosophila* oocytes and nurse cells are roughly the same size, but increase in volume by approximately 40-fold when vitellogenin begins to accumulate about halfway through development of the oocyte. Some vitellogenin is derived from the follicle cells, but most is produced in the fat body and transported to the ovary (Raikhel and Dhadialla 1992). The later stages of oocyte development involve very rapid growth, with the oocyte increasing in volume 1500-fold. While the oocyte is increasing in size, the nurse cells are decreasing because their contents are being deposited in the oocyte. Nurse cells, derived from the germ line, are polyploid, containing 512 and 1024 times the haploid DNA content. These polyploid nurse cells synthesize proteins, ribosomes, and mRNAs. These products, and mitochondria, are transferred to the oocyte by intercellular channels. Thus, the oocyte contains products produced by the mother, which means that initial development in the oocyte is highly dependent upon the genome of the mother (= maternal effects). Finally, the vitelline membrane and the chorion are secreted around the oocyte by follicle cells, and the oocyte enters metaphase of meiosis I. Follicle cells are polyploid, secreting the vitelline membrane of the oocyte and the chorion. The oocyte remains arrested at metaphase of meiosis I until after fertilization.

The oocyte increases in total volume during its development by approximately 90,000-fold. Oogenesis is a complex developmental pathway that is estimated to require the function of 70 to 80% of all genes in the *Drosophila* genome, although the great majority are expressed during other stages of development as well. Only about 75 genes are expressed exclusively during oogenesis (Perrimon et al. 1986). The egg of *D. melanogaster* is rich in stored RNA, including rRNA and mRNA. The bulk of the maternally produced, stored mRNA is derived from transcription of nurse cell nuclei during egg chamber growth, but some mRNA may be derived from the oocyte nucleus itself, which is active briefly about halfway through development. The total amount of mRNA in the oocyte is equal to about 10% of the single-copy DNA of the *Drosophila* genome and corresponds to approximately 8000 distinct protein coding sequences. Most of the mRNA codes for proteins that are required early in embryogenesis, including proteins such as tubulins and histones. Products from a few maternal genes continue to affect development in *D. melanogaster* during the larval stage.

4.13.2. Embryogenesis in *D. melanogaster*

Fertilization occurs when the mature oocytes are released into the oviducts. A single sperm enters the egg cytoplasm through a special channel in the anterior region of the oocyte called the micropyle. Fertilization initiates the completion of meiosis I and II, producing two polar-body nuclei and the female pronucleus. After the haploid male and female pronuclei unite (syngamy), early embryogenesis takes place so rapidly there is no time for cell growth (Figure 4.4). Initial mitoses are atypical because the first nine divisions result in a **syncytium** containing approximately 512 nuclei that lack cellular membranes.

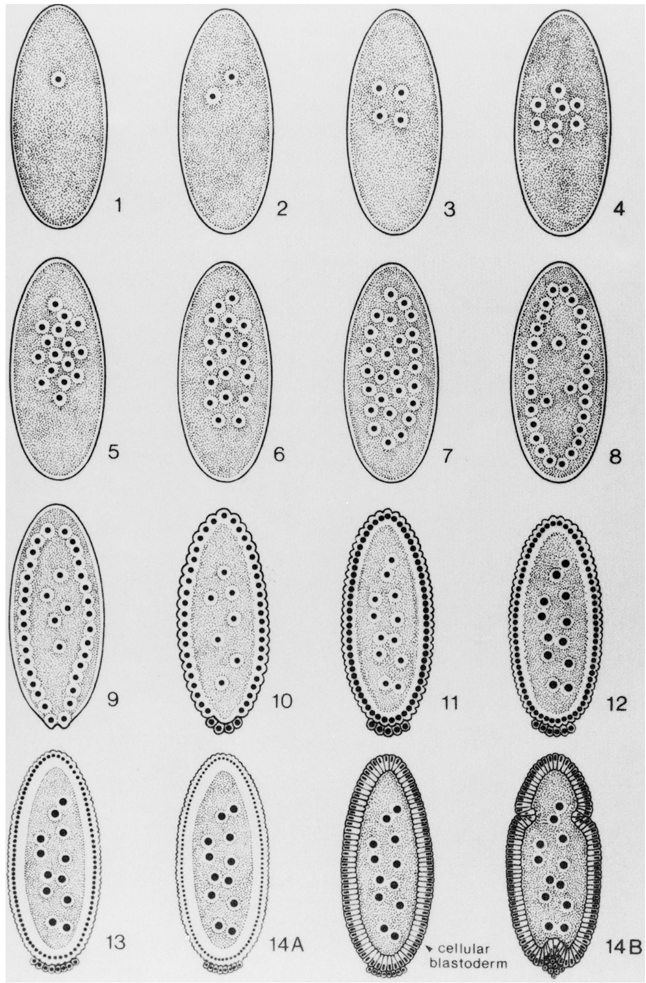


Figure 4.4. Early embryonic stages of *Drosophila melanogaster* from fertilization to just before gastrulation, showing the appearance of pole and somatic buds and cessation of division of yolk nuclei. Numbers indicate division cycles; each cycle begins with the start of interphase and ends at the conclusion of mitosis. Embryos are in longitudinal section without the vitelline membrane. All nuclei (black circles) are shown for cycles 1–5, and afterwards only some are shown. Stippled areas represent yolk, and open areas represent yolk-free cytoplasm. Yolk-free cytoplasm is found both at the periphery (= periplasm) and in islands around the nuclei. During cycles 1–7, nuclei multiply exponentially in the central region of the fertilized egg. Cycle 8 illustrates migration of the majority of the nuclei to the periphery, leaving the future yolk nuclei behind in the center. Yolk nuclei continue to divide in synchrony with other nuclei in cycles 8–10; they then cease dividing and become polyploid. Early in cycle 9, a few nuclei appear in the posterior periplasm and cause protrusions of the cytoplasm, called pole buds. During cycle 10, the remaining migrating nuclei enter the periplasmic region, forming somatic buds over the entire embryonic surface. During the 10th cycle, pole buds are pinched off to form pole cells. After this, synchrony between the pole cells and the syncytium is lost. The syncytial nuclei continue to divide synchronously. The periplasm begins to thicken in cycle 13. During cycle 14, the formation of a plasma membrane begins to separate cells over the entire surface of the embryo, with nuclei elongating to match elongated cells formed by late cycle 14A. During 14B, gastrulation movements begin with the infolding of the cephalic furrow (anterior) and posterior midgut furrow, and subsequently the cells no longer divide synchronously.

After seven nuclear divisions, and when there are 128 nuclei in the central region of the egg, most of the nuclei and their surrounding cytoplasm migrate outward as they continue to divide. A few nuclei are left behind, which divide once to become yolk nuclei that do not become incorporated in the embryo (Figure 4.3). After nine divisions, most of the nuclei have migrated to the egg surface. At this time, the soma and germ-line nuclei segregate when about 15 nuclei move to the posterior region of the egg, bud off, and eventually become the nuclei in the pole cells. These nuclei divide about twice more and become **pole cells** that will give rise to the germ-line tissues of the fly. Meanwhile, the other nuclei migrate to the surface of the egg and divide four times more in synchrony to produce a syncytial blastoderm.

Finally, the membrane covering the egg invaginates to enclose each nucleus in a separate membrane, to form a cellularized **blastoderm** (Figure 4.4). The blastoderm is the layer of cells in an insect embryo that completely surrounds an internal yolk mass. The cellular blastoderm develops from a syncytial blastoderm by partitioning the cleavage nuclei with membranes derived from infolding of the oolemma. During the cellular blastoderm stage, *D. melanogaster* exhibits the **long germ band** type of development in which the pattern of segmentation is established by the end of blastoderm. Some other insects exhibit the **short germ band** type of development in which all or most of the metameric pattern is completed by the sequential addition of segments during elongation of the caudal region of the embryo.

Prior to the cellularized blastoderm stage, the dividing nuclei are equivalent and totipotent, but after the cellularized blastoderm stage is reached, specific body segments have been determined. The cellularized blastoderm stage is a key transition point in embryogenesis in *D. melanogaster* because this is the period during which the products of maternal genes become less important. It is thought that only a few zygotic genes are active prior to cellularization. After the cellularized blastoderm stage, the genes in the zygote begin to dominate in directing the development of the embryo. After additional development, the insect embryo gives rise to a segmented larva with three major tagmata: the head, thorax, and abdomen.

4.13.3. Postembryonic Development

D. melanogaster is a holometabolous insect with sequential life stages: egg → larval stage 1 → molt → larval stage 2 → molt → larval stage 3 → molt → pupa → molt → adult. The larva hatches (= ecloses) from the egg, grows, and molts after each larval stadium. After the third larval stadium, the insect molts, pupates, and undergoes metamorphosis to the adult form. During metamorphosis, most of the larval tissues are digested.

Adult structures develop from cells in structures called **imaginal discs** and abdominal histoblast nests that will give rise to the abdominal epithelium. The cells that give rise to the 19 imaginal discs become segregated from surrounding cells during the first half of embryogenesis. By the time the larva hatches, the imaginal discs and histoblast cells are visibly distinct from the surrounding larval cells because they have smaller nuclei and an undifferentiated appearance. The labial, clypeolabral, antennal + eye, thoracic, three leg, wing, and haltere (= wings on metathorax of other insects) imaginal discs are paired. In addition, there is a single fused genital disc. Imaginal cells are diploid and able to divide. By contrast, most of the 6000 somatic cells of *D. melanogaster* grow in size but do not undergo cell division. The chromosomes of the larval cells continue to undergo replication and become polytene.

At the end of the third larval stage, the larva transforms into a pupa. During the pupal stage, the imaginal discs, each consisting of about 40 cells, develop into adult structures such as legs, wings, eyes, ovaries or testes, and antennae (Larsen-Rapport 1986). Because the imaginal discs were determined during embryonic development, the basic body plan of the adult fly was laid down before the larva eclosed from the egg. The wings, halteres, and legs of the adult, with as many as 50,000 cells each, are formed from the imaginal discs. Not only is segmentation in *Drosophila* based on cues obtained from the mother, but the position and organization of adult structures may also be determined by coordinates provided by the mother (Couso and Gonzalez-Gaitan 1993). After emergence as an adult, the insect mates and the progeny begin this developmental cycle again.

4.14. Dissecting Development with *D. melanogaster* Mutants

The study of development in *Drosophila* is dependent upon the availability of mutants so that the process can be dissected. In fact, in discussing development in *Drosophila*, the genes influencing development are called by names that reflect their mutant form. Nusslein-Volhard and Wieschaus (1980) began a systematic program of mutagenizing *Drosophila* females in order to obtain many developmental mutants in insect embryos. In addition, many useful developmental mutants were discovered by E. B. Lewis in his pioneering work on *Drosophila* development involving the bithorax homeotic gene complex. Nusslein-Volhard, Wieschaus, and Lewis jointly were awarded the 1995 Nobel Prize in Medicine for their pioneering research on development.

The term **homeotic** was coined to describe the replacement of one part of the body by a serially homologous part. Lewis (1978) developed the hypothesis that families of structurally related genes control the specification of the insect body plan because insect bodies are metameric (composed of serially repeating units or body segments) which differentiate into specific structures according to their position. Likewise, the appendages in each major body segment develop into appropriate structures, with antennae located on the head, legs on the thoracic segments, and wings on the second thoracic segment. However, this normal pattern is disrupted in a number of homeotic mutants.

Drosophila embryos appear to go through two phases in their development. During the first phase, many genes appear to encode transcription factors or nuclear proteins. This suggests that development is regulated by a cascade of transcriptional factors that regulate other genes. There is a successive subdivision of the embryo into smaller and smaller domains that is accomplished by the differential and combinatorial action of transcription factors. The first phase is completed by the time cells are fully formed at the end of the blastoderm stage.

The second phase begins after the formation of the cellular blastoderm and consists of elaborating the information provided from reference points that have been deposited along the dorsal–ventral and anterior–posterior axes. This requires the communication of information between cells by intercellular signal molecules.

The genes that control *Drosophila* embryonic development can be divided into three classes: 1) **maternal-effect genes** that specify egg polarity and the spatial coordinates of the egg and future embryo, 2) **segmentation genes**, including the gap, pair-rule, and segment polarity classes of genes, that determine the number and polarity of the body segments, and 3) **homeotic genes** that determine the identification and sequence of the

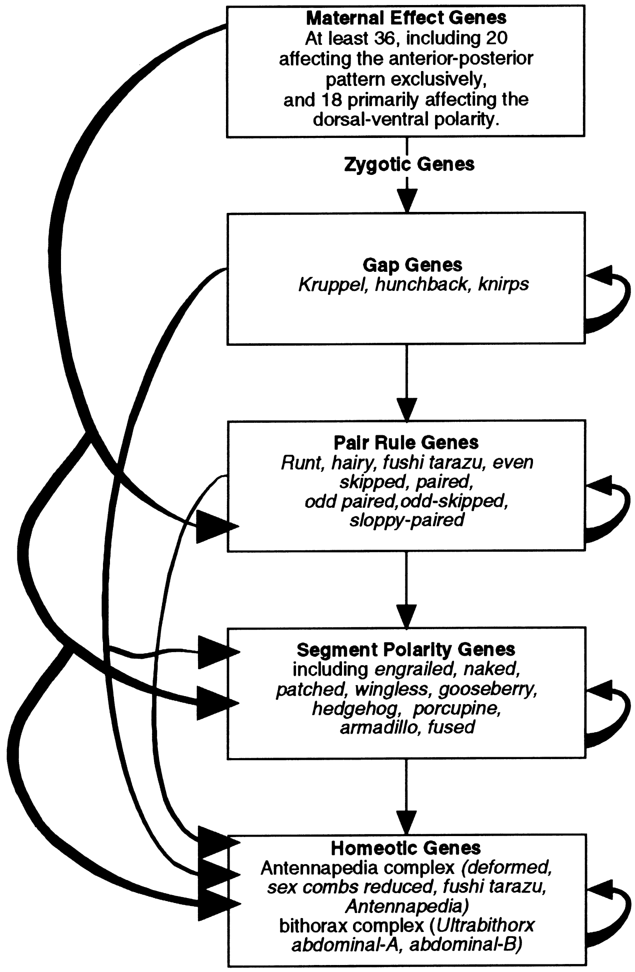


Figure 4.5. Development of segments in the embryo of *Drosophila melanogaster* involves a hierarchy of regulatory genes. The earliest acting genes are maternal-effect genes that regulate the spatial expression of later-acting genes. Some maternal gene products are present in concentration gradients and regulate downstream genes in a concentration-dependent manner. In many cases, genes within a given class regulate other members of the class (depicted as arrows beginning and ending at the same class). Many other genes not shown may also play a role in segmentation.

segments. Although most genes with a homeo domain are in the homeotic class, a few are found among the segmentation genes (Figure 4.5).

4.14.1. Maternal-Effect Genes

Maternal-effect genes function in the mother, and their products (mRNA or stored materials) influence development of oocytes and embryos. Maternal-effect genes are discovered by determining if the mother carries a mutant gene and her embryo cannot be rescued by a wild-type gene contributed by the father at the time the egg is fertilized. Maternal-effect

genes are important because the basic organization of the oocyte has been accomplished even before the egg has been fertilized. At least 38 maternal-effect genes have been studied; about 20 affect the anterior–posterior pattern, and about 18 affect the dorsal–ventral polarity of the embryo (Figure 4.5).

The 20 genes influencing differentiation into proper anterior–posterior polarity can be subdivided into a group affecting the anterior half of the embryo (including *bicaudal*$^+$, *Bicaudal C*$^+$, *Bicaudal D*$^+$, *bicoid*$^+$, *exuperantia*$^+$, *swallow*$^+$), a group affecting the posterior half (including *torso*$^+$, *trunk*$^+$, *fs (1) polehole*$^+$, *fs (1) nasrat*$^+$, *lethal (1) polehole*$^+$), and a group affecting both the anterior and posterior ends (including *oskar*$^+$, *staufen*$^+$, *tudor*$^+$, *valois*$^+$, *vasa*$^+$, *cappuccino*$^+$, *spire*$^+$, *nanos*$^+$, and *pumilio*$^+$). Of the 18 genes affecting dorsoventral polarity, several are required for the polarity of both the eggshell and embryo, and several are required for the polarity of the embryo only (Wilkins 1993).

Determining the dorsal–ventral and anterior–posterior polarity in the embryo is a highly significant step. The *Toll*$^+$ gene is a dorsalizing gene, and mutations produce embryos lacking both ventral and lateral structures. The *Toll*$^+$ gene product appears be a **morphogen** (molecules whose local concentration directly *determine* the local pattern of differentiation), because when the wild-type gene product is injected into mutant embryos it can make the affected region become the dorsal region of the fly.

A major determinant of anterior–posterior polarity is the product of the *bicoid*$^+$ gene. It is transcribed in the nurse cells of the ovary, and the mRNA passes into the oocyte where it becomes localized in the anterior of the egg, apparently aided by components of the cytoskeleton that are encoded by products of the genes *swallow*$^+$ and *exuperentia*$^+$.

Several of the maternal-effect genes (*nanos*$^+$, *cappuccino*$^+$, *spire*$^+$, *staufen*$^+$, *oskar*$^+$, *vasa*$^+$, *valois*$^+$, and *tudor*$^+$) are required for the localization of factors that determine the germ line. In addition, mitochondrial large ribosomal RNA (mtlrRNA) may be important for pole cell formation (Kobayashi et al. 1993). Thus, both mitochondrial and nuclear genomes are involved in determination of the germ line during embryonic development.

Maternal-effect genes are most important during development of the egg up to the blastoderm stage. After that, genes inherited by the zygote from both parents become dominant factors determining development. However, because development is an elaborative process and the adult phenotype is a summation of the developmental effects accrued over the life span of the individual, developmental events early in the life cycle can significantly influence the phenotype at later stages. For example, maternal effects have significant and diverse effects on insect life histories, including incidence and intensity of diapause, production of sexual forms, wing polyphenism, dispersal behavior, development time, growth rate, resistance to chemicals and microbial infection, and survival. Some of these influences are caused by maternal age and diet, but some are genetically determined (Mousseau and Dingle 1991).

4.14.2. Zygotic Segmentation Genes

During blastoderm formation, the embryo begins to develop a pattern of repeating body segments. The genetic control of segmentation is determined by zygotic genes that have been divided into three categories: 1) pair-rule, 2) gap, and 3) segment polarity genes.

Segmentation mutants found in *Drosophila* embryos initially were difficult to interpret because they did not affect what appeared to be a single segment; usually they affected half of one "segment" and the adjacent half of the next. Eventually, it was determined that

true segments are not reflected by the visible cuticular patterns of sclerites and sutures; visible segments are, in fact, **parasegments**. There are 14 complete parasegments in *D. melanogaster* that are defined early in development; each is a precise set of cells.

Gap-gene mutants cause deletions in groups of adjacent segments, pair-rule mutants cause pattern deletions in alternating segments, and segment-polarity mutants cause pattern defects in every segment (Figure 4.5). Most segmentation mutants are lethals in the zygote, but some gap genes have a maternal effect and are expressed during oogenesis. Four of the segmentation genes (*fushi tarazu*$^+$, *even-skipped*$^+$, *paired*$^+$, and *engrailed*$^+$) contain a homeobox (see below). Thus, these genes encode DNA regulatory proteins or transcription factors that bind to specific DNA or RNA sequences.

4.14.2.1. Gap Genes

Gap genes were so named because large areas of the normal cuticular pattern are deleted in individuals with mutant phenotypes (Figure 4.6). The three wild-type versions of the gap genes, *Krüppel*$^+$, *hunchback*$^+$, and *knirps*$^+$, regionalize the embryo by delimiting domains of homeotic gene expression and affect position-specific regulation of the pair-rule genes (*runt*$^+$, *fushi tarazu*$^+$, *even skipped*$^+$, *paired*$^+$, and *odd-paired*$^+$). All three gap-gene products contain DNA-binding domains. Embryos with the mutated version of *Krüppel* lack all of the thoracic and most of the abdominal segments, while *knirps* embryos have a normal thoracic region but lack nearly all the abdominal segments. Embryos homozygous for *hunchback* lack head segments, mesothorax, and metathorax while showing a normal abdominal segmentation.

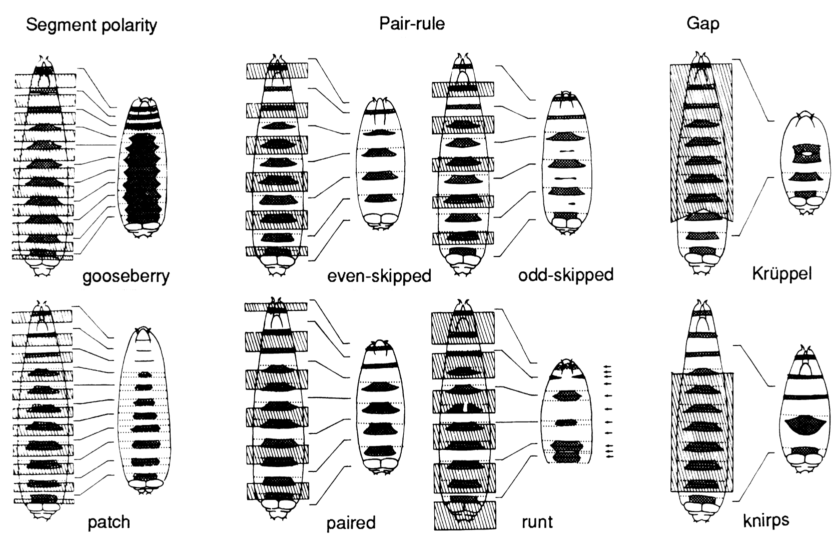


Figure 4.6. Embryonic segment pattern defects are illustrated with selected mutants: *gooseberry* and *patch* are segment polarity mutants; *even-skipped*, *odd-skipped*, *paired*, and *runt* are pair-rule mutants; and *Krüppel*$^+$ and *knirps*$^+$ are gap genes. Dotted regions represent denticle bands on the developing embryo, and dotted lines show the boundaries of segments. Hatched regions indicate the parts of the pattern that are missing in the mutants. Transverse lines link the corresponding regions in mutant and wild-type embryos. Arrows indicate where lines of polarity have been reversed. (From *Nature* 287: 796, 1980.)

The gap genes interact to produce sharp boundaries, with wild-type alleles of *hunchback*$^+$ and *Krüppel*$^+$ repressing one another and *knirps*$^+$ acting as a negative regulator of *Krüppel*$^+$. Thus, the establishment of stable domains by the gap genes is a two-step process: 1) a differential response to graded levels of maternal determinants, 2) followed by mutual repression leading to the generation of stable boundaries between adjacent domains. Gap genes regionalize the embryo by delimiting the domains of later homeotic gene expression, and this results in position-specific regulation of the pair-rule class of genes, which give rise to the metamerization of the embryo.

4.14.2.2. Pair-Rule Genes

The **pair-rule genes** were so named because mutant flies have a repetitive aberration throughout the germ band, with the removal of integral, alternate segment-width areas (Figure 4.6). The pair-rule genes (including *runt*$^+$, *hairy*$^+$, *fushi tarazu*$^+$, *even skipped*$^+$, *paired*$^+$, *odd-paired*$^+$, *odd-skipped*$^+$, *sloppy-paired*$^+$) are transiently expressed in seven or eight stripes during cellularization of the blastoderm. However, each is otherwise unique in its expression. Both *runt*$^+$ and *hairy*$^+$ are initially expressed rather uniformly throughout the embryo but begin restricted expression earlier than the others in the class. Because *runt*$^+$ and *hairy*$^+$ appear to have a major role in generating the striped pattern of the other pair-rule genes, they are called the primary pair-rule genes. The name *fushi tarazu*$^+$ (*ftz*$^+$) is a Japanese term meaning "not enough segments," and mutants of this pair-rule gene have half the normal number of segments. Pair-rule genes are essential, directly or indirectly, for the initial establishment of segmentation.

4.14.2.3. Segment Polarity Genes

The **segment polarity genes** appear to determine a linear sequence of repeated positional values within each segment. The segment polarity mutants have a repetitive deletion of pattern, but the deletions occur within each segment and are followed, for many mutants in this group, by a partial mirror-image duplication of the part that remains (Figure 4.6). Segment polarity genes (including *engrailed*$^+$, *naked*$^+$, *patched*$^+$, *wingless*$^+$, *gooseberry*$^+$, *patched hedgehog*$^+$, *porcupine*$^+$, *armadillo*$^+$, *fused*$^+$) are required either continuously or over extensive periods to maintain the segmentation pattern. Most or all are required to maintain patterns in the imaginal tissues.

4.14.2.4. Homeotic Genes

In 1894 William Bateson coined the word **homeosis** to describe the situation in which "something has been changed into the likeness of something else" (Lewis 1994). Bateson was attempting to provide evidence in support of Darwin's theory of evolution, and homeotic variations appeared to Bateson to be the kind of dramatic changes that could explain how evolution occurred. E. B. Lewis (1994) concluded that homeosis provided a rich legacy: "Besides giving us the homeobox, it has opened up a completely new approach to the study of development. And over the past 15 years, it has led to the realization that the body plan of most animals, and presumably of plants as well, is controlled by a set of master regulatory genes, first identified by their homeotic mutations."

The periodic pattern of body segments generated by segmentation genes (gap genes, pair-rule genes, and segment polarity genes) has to be converted into segments with

different characteristics (Figure 4.5). Thus, in insects, thoracic segment 2 is different from thoracic segment 3 and abdominal segment 2 will be different from the terminal abdominal segments, which typically have genital structures. This fine tuning is determined by **homeotic genes**. Both homeotic and some segmentation genes contain a special sequence called the **homeobox**. The homeobox consists of about 180 bp, which is found in the 3′ exon of the *Ultrabithorax*$^+$, *abdominal-A*$^+$, *Abdominal-B*$^+$, *Deformed*$^+$, *sex combs reduced*$^+$, *fushi tarazu*$^+$, and *Antennapedia*$^+$ genes. The sequences of the different homeoboxes are nearly identical and they mediate the binding of homeotic proteins to specific DNA sequences and thus regulate the expression of many downstream genes. It has been proposed that just two homeotic genes, *even-skipped*$^+$ and *fushi tarazu*$^+$, directly control the expression of the majority of genes in the *Drosophila* genome (Mannervik 1999). Homeodomain proteins occur in all eukaryotes, where they perform important functions during development.

Since the first homeobox sequence was isolated from the *Antennapedia*$^+$ gene in late 1983, it has been used as a probe to identify and isolate previously unknown homeotic genes from *Drosophila*. Furthermore, because the homeobox is evolutionarily conserved, this *Drosophila* sequence was used as a probe to identify homeotic genes from other species, including humans (Gehring 1985). The homeobox sequence codes for 60 amino acids, and its presence labels a protein as a DNA binding protein and the gene containing it as one that can control other genes. One of the homeobox sequences of *Xenopus* is so similar to the *Antennapedia* homeobox in *Drosophila* that only one amino acid out of 60 is different. The reason for this extraordinary conservation during evolution is not fully understood (Gehring 1987). However, not only are the sequences of the different elements in the vertebrate and insect gene clusters comparable, but the order of those elements on the chromosome is conserved. Thus, the position of the anterior margin of expression in the body axis (*Ultrabithorax*$^+$ affects parasegment 5; *abdominal-A*$^+$ affects parasegment 7; *Abdominal-B*$^+$ affects parasegment 10) and their order on the chromosome correlate. The order of the homologous homeobox genes in vertebrates also is conserved. This remarkable conservation suggests that there may have been an ancestral sequence common to flies and humans that survives in both the hindbrain of humans and the parasegments of insects (Lawrence 1992).

Homeotic mutants may have segments that are transformed dramatically. For example, antennal segments may be transformed into leglike structures, and metathoracic segments with halteres may be transformed into mesothoracic segments with a set of wings (Figure 4.7). The four-winged *D. melanogaster* shown in Figure 4.7 is the result of combining three separate mutated genes in one fly! Normally, of course, a pair of wings is found on the second thoracic segment, and a small pair of balancing organs, called halteres, are found on the third thoracic segment. However, this fly has two essentially normal second thoraces (and no third thoracic segment) because the combined effect of the three mutations is to transform the third thoracic segment into the second without affecting any other parts of the fly.

E. B. Lewis (1978) proposed a combinatorial model that assumes each insect segment is specified by a unique combination of homeotic genes that are expressed in that particular segment. Thus, the fewest homeotic genes would be required in thoracic segment 2, which would be the prototypical segment, and progressively more genes would be active in the more posterior segments. Although this model has been modified, it provided a useful conceptual framework for investigating *Drosophila* development.

Homeotic genes have some unusual characteristics. First, several homeotic genes appear to be very large relative to most other genes in *Drosophila*. For example, the *Antennapedia*$^+$



Figure 4.7. A four-winged *Drosophila melanogaster* fly showing a complete transformation of the third thoracic segment into a second thoracic segment. The fly carries one chromosome with a deletion of the homeotic bithorax complex, while the other chromosome carries mutations of the *bithorax* (*Abx*, *bx3*) and *postbithorax* (*pbx*) loci. (Photograph provided by Edward Lewis, California Institute of Technology.)

primary gene transcript is approximately 100 kb long and the *Ultrabithorax*$^+$ transcript is about 75 kb long. However, after the introns are spliced out, the remaining sequences are only a few kilobases long. Many of the exons in homeotic genes appear to encode protein domains with distinct structural or enzymatic functions. As a result, alternative splicing patterns in large genes such as the Antennapedia and bithorax complexes may allow organisms to adapt one basic protein structure to different, but related, developmental uses. By adding or subtracting functional protein domains encoded by optional exons, the structural and enzymatic properties of the homeotic gene product can be modified and the ability of the protein to interact with other cellular components can be altered as development proceeds.

4.15. Interactions during Development

Normal development requires the coordinated expression of thousands of structural genes in a controlled manner. Because independent control of individual structural genes would result in chaotic development, controlling genes regulate the activity of groups of structural genes in a coordinate manner. Such genes are presumably arranged hierarchically or form a controlling network that ensures the proper timing of development so that the proper pattern develops. Although development in *Drosophila* is not fully understood, many of the genes and their interactions are known, including those that regulate the development of mesoderm (Furlong et al. 2001), appendages (Morata 2001), and the eye (Thomas and Wassarman 1999). Progress toward understanding the full process is rapid. Furthermore, it is likely that understanding development in *D. melanogaster* will elucidate many of the principles by which other higher eukaryotes develop. The molecular genetics of sex determination is discussed in Chapter 10 and provides another example of the hierarchical nature of development.

4.16. Similarities and Differences in Development in Other Insects

Although developmental studies are most advanced in *D. melanogaster*, it is important to carry out comparable research with other insects to determine whether our knowledge of development in *D. melanogaster* can be extrapolated to other insects (Klingler 1994, Buning 1994). *Drosophila* represents a relatively specialized type of development, the "long germ" development pattern, in which segmentation occurs essentially simultaneously along the anterior–posterior axis with the process of segmental specification under the control of homeotic genes in the Antennapedia and bithorax complexes. Research on development in other insects includes studies of the locust (Orthoptera) *Schistocerca gregaria*, the house fly *Musca domestica*, the honey bee *Apis mellifera*, the silkworm *Bombyx mori*, the moth-midge *Clogmia albipunctata* (Diptera), and Thysanura (firebrats), as well as the flour beetle *Tribolium castaneum*.

4.16.1. Development in *Tribolium*

Beeman (1987) showed that six loci of homeotic genes in a single cluster (HOM-C) of *T. castaneum* contain elements homologous to the homeotic genes in the Antennapedia and bithorax complexes of *Drosophila*. These genes map along the chromosome in the same order from anterior to posterior as their effects occur, but they occur in a single cluster rather than in two as in *Drosophila* and with a different gene order (Beeman et al. 1989, 1993a,b). In mammals, there are four clusters of homeotic genes that are homologous to these genes, and the gene order in both insects and vertebrates is similar. This suggests there is a common ancestral gene cluster for both insects and vertebrates (Beeman et al. 1993b).

Tribolium with the *maxillopedia* mutation have labial and maxillary palps transformed into leglike structures (Figure 4.8B). *Tribolium* with mutations of *Cephalothorax*$^+$ have the head and first thoracic segment fused and the labial palps transformed into antennae (Figure 4.8D). *Tribolium* with *prothoraxless*$^+$ mutations exhibit fusion of head segments with the entire thorax, and transformation of all three pairs of thoracic legs into antennae.

Despite unique components and differences between *D. melanogaster* and other arthropods, analyses suggest that *D. melanogaster* serves as a useful model for understanding development in arthropods. Comparative studies on the molecular evolution of genes involved in development in insects and other organisms thus may provide information about the evolution of gene families regulating development, as well as contribute to our understanding of the basic mechanisms underlying the genetic control of development.

4.17. Evo-Devo and the Revolution in Developmental Studies

A new discipline of "evolutionary developmental biology," also called "evo-devo," has recently emerged in which the diversity of development in animal and plant forms is studied from an evolutionary point of view (Pennisi and Roush 1997, West-Eberhard 1998, Holland 1999, von Dassow and Munro 1999, Dalton 2000, Jenner 2000). Evo-devo combines fields as diverse as comparative embryology, paleontology, molecular phylogenetics, and genome

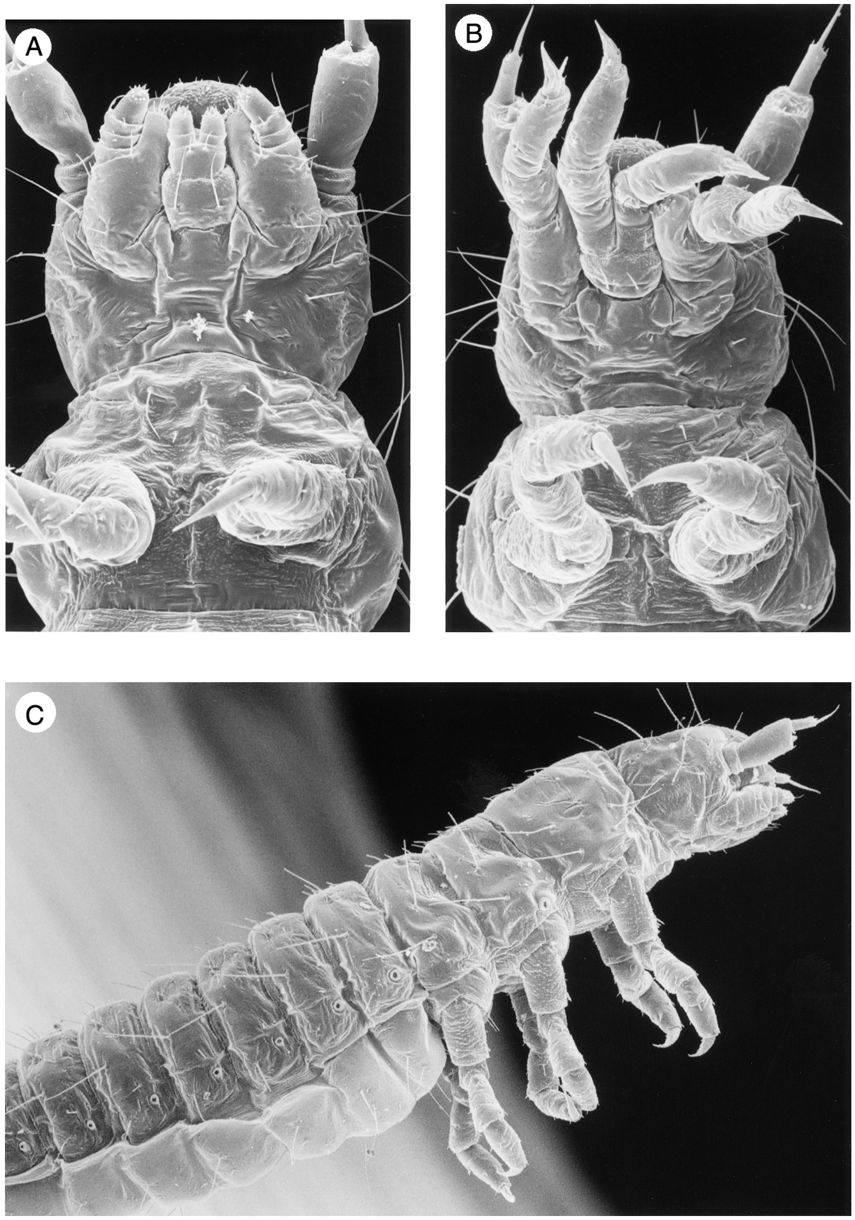


Figure 4.8. A) A ventral view of a wild-type embryo of *Tribolium castaneum*. B) A ventral view showing a homeotic mutation, *maxillopedia*, in which the maxillary and labial palps are modified into leg-like appendages. C) A side view of a mature wild-type embryo. D) A side view showing the homeotic mutant *cephalothorax*, which results in incorporation of the prothorax into the head and transformation of the labial palps into antennae. *Maxillopedia* and *cephalothorax* are in the HOM-C complex and correspond to *proboscipedia* and sex combs reduced in *Drosophila*. (Photographs provided by Richard W. Beeman.)



Figure 4.8. *continued*

analysis. The goals of **evo-devo** (evolutionary developmental biology) are to understand (Hall 2000):

- The origin and evolution of embryonic development
- How modifications of development and developmental processes lead to the production of novel features
- The adaptive plasticity of development in life-history evolution
- How ecology affects development to modulate evolutionary change
- The developmental basis of homoplasy (parallel or convergent evolution) and homology

Research in evo-devo is undertaken with the understanding that the gene is the hereditary unit and that development is hierarchical and characterized by emergent properties whose features cannot be predicted from properties at a lower level in the hierarchy. The assumption underlying evo-devo is that analysis of the evolution of developmental stages, processes, and mechanisms will enable us to understand how organisms, organs, tissues, cells, and genes evolve. Evo-devo successes include learning that (Hall 2000):

- Genes that control major developmental processes (establishment of body plans, formation of appendages and sense organs) are shared across the animal kingdom and arose early in metazoan evolution.
- New knowledge of developmental mechanisms underlying the formation of organs or major body parts has led to an understanding of the mechanisms involved in their origin from structures in ancestral organisms.
- Loss of organs does not imply loss of the developmental potential to form those organs.
- Life-history stages (embryos, larvae, adults) can develop and evolve separately, which provides opportunities to modify and modulate embryonic development, for specialization or diversification of adult structure, and for the evolution of novel structures.

- Homology is now seen as hierarchical, with homologous genes initiating development of structures that are not homologous (such as arthropod and vertebrate eyes) and homologous structures developing by processes that are not homologous.

For example, studies seem to confirm that there was an inversion of the dorsoventral axis during animal evolution, with the ventral region of *Drosophila* homologous to the dorsal side of vertebrates (De Robertis and Sasai 1996). Thus, developmental systems that control patterns from eggs to adults are remarkably similar across a wide range of phyla despite at least half a billion years of evolution since their origin from their last common ancestor (Erwin et al. 1997). Despite their genetic similarities, the developmental systems produce very different body plans. The basic aspects of the developmental control systems of long-extinct animals can now be reconstructed, and the diversification of animal form and the evolution of the genetic controls that regulate it are becoming understood (Erwin et al. 1997).

The field of evo-devo uses the concept of homology. However, **homology** is a difficult and, sometimes, fuzzy term (Bolker and Raff 1996). There are at least nine homology concepts in current use in the comparative biology literature, which makes the developing field of evo-devo a particularly contentious one (Janies and DeSalle 1999). The concept of homology originated from classical studies of comparative adult morphology, and one classical definition is "the same organ in different animals under every variety of form and function." Another is "derived from an equivalent characteristic of the common ancestor." Homology has been used recently to mean "shared patterns of gene expression," a controversial use of the term (Bolker and Raff 1996) and more often is limited to "similarities due to descent from the same ancestral source." Unfortunately, most genes play multiple roles in development, making it difficult to resolve homologies. For example, arthropods, annelids, and chordates all possess segments. It remains unclear whether the segments evolved independently or were derived from a common ancestor (Davis and Patel 1999).

New technology allows new approaches to understanding development, and the development of vast amounts of DNA sequence information allows researchers to determine which genes are functioning from a global perspective (White et al. 1999, Janies and DeSalle 1999). For example, it is possible, using microarray analysis (described in Chapter 7), to evaluate the role of hundreds of genes involved in the process of metamorphosis of *Drosophila*. The field of evo-devo promises to enliven discussions of development.

References Cited

Abe, H., F. Ohbayashi, T. Shimada, T. Sugasaki, S. Kawai, and T. Oshiki (1998). A complete full-length non-LTR retrotransposon BMC1, on the W chromosome of the silkworm, *Bombyx mori*. *Genes Genet. Syst.* **73**: 353–358.

Abe, H., F. Ohbayashi, T. Shimada, T. Sugasaki, S. Kawai, K. Mita, and T. Oshiki (2000). Molecular structure of a novel *gypsy*-Ty3-like retrotransposon (*Kabuki*) and nested retrotransposable elements on the W chromosome of the silkworm *Bombyx mori*. *Mol. Gen. Genet.* **263**: 916–924.

Abe, H., F. Ohbayashi, T. Sugasaki, M. Kanehara, T. Terada, T. Shimada, S. Kawai, K. Mita, Y. Kanamori, M. T. Yamamoto, and T. Oshiki (2001). Two novel *Pao*-like retrotransposons (*Kamikaze* and *Yamato*) from the silkworm species *Bombyx mori* and *B. mandarina*: common structural features of *Pao*-like elements. *Mol. Genet. Genomics* **265**: 375–385.

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Agarwal, M., N. Bensaadi, J. C. Salvado, K. Campbell, and C. Mouches (1993). Characterization and genetic organization of full-length copies of a LINE retroposon family dispersed in the genome of *Culex pipiens* mosquitoes. *Insect Biochem. Mol. Biol.* **23**: 621–629.

Aksoy, S. (2000). Tsetse—a haven for microorganisms. *Parasitology Today* **16**: 114–118.

Anonymous. (2001). Gene sequencers hope to put the bite on mosquitoes. *Nature* **410**: 137.

Antoine, M., and J. Niessing (1984). Intron-less globin genes in the insect *Chironomus thummi thummi*. *Nature* **310**: 795–798.

Ashburner, M. (1989). *Drosophila*. *A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bachtrog, D., S. Weiss, B. Zangerl, G. Brem, and C. Schlotterer. 1999. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol. Biol. Evol.* **16**: 602–610.

Balter, M. (2001). Sequencing set for dreaded mosquito. *Science* **291**: 1873.

Bandi, C., B. Slatko, and S. L. O'Neill (1999). *Wolbachia* genomes and the many faces of symbiosis. Parasitol. *Today* **15**: 428–429.

Baumann, P., N. A. Moran, and L. Baumann (1997). The evolution and genetics of aphid endosymbionts. *BioScience* **47**: 12–20.

Bazzocchi, C., W. Jamnongluk, S. L. O'Neil, T. J. C. Anderson, C. Genchi, and C. Bandi (2000). *wsp* gene sequences from the *Wolbachia* of filarial nematodes. *Curr. Microbiol.* **41**: 96–100.

Beckage, N. E. (1998). Parasitoids and polydnaviruses. *BioScience* **48**: 305–310.

Beeman, R. W. (1987). A homeotic gene cluster in the red flour beetle. *Nature* **327**: 247–249.

Beeman, R. W., J. J. Stuart, M. S. Haas, and R. E. Denell (1989). Genetic analysis of the homeotic gene complex (Hom-C) in the beetle *Tribolium castaneum*. *Dev. Biol.* **133**: 196–209.

Beeman, R. W., J. J. Stuart, S. J. Brown, and R. E. Denell (1993a). Structure and function of the homeotic gene complex (HOM-C) in the beetle, *Tribolium castaneum*. *BioEssays* **15**: 439–444.

Beeman, R. W., S. J. Brown, J. J. Stuart, and R. Denell (1993b). The evolution of genes regulating developmental commitments in insects and other animals, pp. 71–84. In: *Evolutionary Conservation of Developmental Mechanisms.* Wiley-Liss, New York.

Beeman, R. W., M. S. Thomson, J. M. Clark, M. A. DeCamillis, S. J. Brown, and R. E. Denell (1996). *Woot*, an active Gypsy-class retrotransposon in the flour beetle, *Tribolium castaneum*, is associated with a recent mutation. *Genetics* **143**: 417–426.

Bello, B., and P. Couble (1990). Specific expression of a silk-encoding gene of *Bombyx* in the anterior salivary gland of *Drosophila*. *Nature* **346**: 480–482.

Bensaadi-Merchermek, N., C. Cagnon, I. Desmons, J. C. Salvado, S. Karama, F. D'Amico, and C. Mouches (1997). CM-gag, a transposable-like element reiterated in the genome of *Culex pipiens* mosquitoes, contains only a *gag* gene. *Genetica* **100**: 141–148.

Berg, D. E., and M. M. Howe, Eds. (1989). *Mobile DNA*. Am. Soc. Microbiol., Washington, DC.

Berry, S. J. (1985). Insect nucleic acids, pp. 219–253. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology*. Vol. (10). G. A. Kerkut, and L. I. Gilbert, Eds. Pergamon Press, Oxford, UK.

Besansky, N. J. (1990a). A retrotransposable element from the mosquito *Anopheles gambiae*. *Mol. Cell. Biol.* **10**: 863–871.

Besansky, N. J. (1990b). Evolution of the T1 retroposon family in the *Anopheles gambiae* complex. *Mol. Biol. Evol.* **7**: 229–246.

Besansky, N. J., and J. R. Powell (1992). Reassociation kinetics of *Anopheles gambiae* (Diptera: Culicidae) DNA. *J. Med. Entomol.* **29**: 125–128.

Besansky, N. J., J. A. Bedell, and O. Mukabayire (1994). *Q*: a new retrotransposon from the mosquito *Anopheles gambiae*. *Insect Mol. Biol.* **3**: 49–56.

Beye, M., A. Poch, C. Burgtorf, R. F. A. Moritz, and H. Lehrach (1998). A gridded genomic library of the honeybee (*Apis mellifera*): a reference library system for basic and comparative genetic studies of a hymenopteran genome. *Genomics* **49**: 317–320.

Biemont, C., C. Vieira, N. Borie, and D. Lepetit (1999). Transposable elements and genome evolution: the case of *Drosophila simulans*. *Genetica* **107**: 113–120.

Biessmann, H., B. Kasravi, K. Jakes, T. Bui, K. Ikenaga, and J. M. Mason (1993). The genomic organization of HeT-A retroposons in *Drosophila melanogaster*. *Chromosoma* **102**: 297–305.

Biessmann, H., M. F. Walter, D. Le, S. Chuan, and J. G. Yao (1999). *Moose*, a new family of LTR-retrotransposons in the mosquito *Anopheles gambiae*. *Insect Mol. Biol.* **8**: 201–212.

Bigot, Y., F. Lutcher, M. H. Hamelin, and G. Periquet (1992). The 28S ribosomal RNA-encoding gene of Hymenoptera: inserted sequences in the retrotransposon-rich regions. *Gene* **121**: 347–352.

Blinov, A. G., Y. V. Sobanov, S. V. Scherbik, and K. G. Aimanova (1997). The *Chironomus* (*Camptochironomus*) genome contains two non-LTR retrotransposons. *Genome* **40**: 143–150.

Block, K., G. Ising, and F. Stahl (1990). Minichromosomes in *Drosophila melanogaster* derived from the transposing element TE1. *Chromosoma* **99**: 336–343.

Bolker, J. A., and R. A. Raff (1996). Developmental genetics and traditional homology. *BioEssays* **18**: 489–494.

Booth, D. R., P. D. Ready, and D. F. Smith 1994. Isolation of non-LTR retrotransposon reverse transcriptase-like sequences from phlebotomine sandflies. *Insect Mol. Biol.* **3**: 89–96.

Booth, D. R., P. D. Ready, and D. F. Smith (1996). Evolution of multiple families of non-LTR retrotransposons in phlebotomine sandflies. *Genet. Res. Camb.* **67**: 227–237.

Bourtzis, K., S. L. Dobson, H. R. Braig, and S. L. O'Neill (1998). Rescuing *Wolbachia* have been overlooked. *Nature* **391**: 852–853.

Bowen, N. J., and J. F. McDonald (2001). *Drosophila* euchromatic LTR-retrotransposons are much younger than the host species in which they reside. *Genome Res.* **11**: 1527–1540.

Bownes, M. (1986). Expression of the genes coding for vitellogenin (yolk protein). *Annu. Rev. Entomol.* **31**: 507–531.

Boyle, L. S., L. O'Neill, H. M. Robertson, and T. L. Karr (1993). Interspecific and intraspecific horizontal transfer of *Wolbachia* in *Drosophila*. *Science* **260**: 1796–1799.

Britten, R. J. (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**: 177–182.

Brody, T. (1999). The Interactive Fly: gene networks, development and the Internet. *Trends Genet.* **15**: 333–334.

Brosius, J. (1991). Retroposons—seeds of evolution. *Science* **251**: 753.

Brown, S. W., and U. Nur (1964). Heterochromatic chromosomes in coccids. *Science* **145**: 130–136.

Brown, S. J., J. K. Henry, W. C. Black IV, and R. E. Denell (1990). Molecular genetic manipulation of the red flour beetle: genome organization and cloning of a ribosomal protein gene. *Insect Biochem.* **20**: 185–193.

Brune, A., and M. Friedrich (2000). Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**: 263–269.

Buning, J. (1994). *The Insect Ovary. Ultrastructure, Previtellogenis Growth, and Evolution*. Chapman and Hall, New York.

Burke, W. D., D. G. Eickbush, Y. Xiong, J. Jakubczak, and T. H. Eickbush (1993). Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. *Mol. Biol. Evol.* **10**: 163–185.

Calvi, B. R., T. J. Hong, S. D. Findley, and W. M. Gelbart (1991). Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *activator*, and *Tam3*. *Cell* **66**: 465–471.

Carminati, J. L., C. G. Johnston, and T. L. Orr-Weaver (1992). The *Drosophila* ACE3 chorion element autonomously induces amplification. *Mol. Cell. Biol.* **12**: 2444–2453.

Carton, Y., and A. J. Nappi (2001). Immunogenetic aspects of the cellular immune response of *Drosophila* against parasitoids. *Immunogenetics* **52**: 157–164.

Cazemier, A. E., J. H. P. Hackstein, H. J. M. Op den Camp, J. Rosenberg, and C. van der Drift (1997). Bacteria in the intestinal tract of different species of arthropods. *Microb. Ecol.* **33**: 189–197.

Chapco, W. (1983). Genetics and population genetics of grasshoppers and locusts—a bibliography. *Bibliog. Entomol. Soc. Am.* **2**: 11–16.

Coates, C. J., K. N. Johnson, H. D. Perkins, A. J. A. Howells, D. A. O'Brochta, and P. W. Atkinson (1996). The *hermit* transposable element of the Australian sheep blowfly, *Lucilia cuprina*, belongs to the *hAT* family of transposable elements. *Genetica* **97**: 23–31.

Cook, J. M., and R. D. J. Butcher (1999). The transmission and effects of *Wolbachia* bacteria in parasitoids. *Res. Popul. Ecol.* **41**: 15–28.

Cook, J. M., J. Martin, A. Lewin, R. E. Sinden, and M. Tristem (2000). Systematic screening of *Anopheles* mosquito genomes yields evidence for a major clade of *Pao*-like retrotransposons. *Insect Mol. Biol.* **9**: 109–117.

Couso, J. P., and M. Gonzalez-Gaitan (1993). Embryonic limb development in *Drosophila*. *Trends Genet.* **9**: 371–373.

Craig, C. L. (1997). Evolution of arthropod silks. *Annu. Rev. Entomol.* **42**: 231–267.

Crozier, R. H., and Y. C. Crozier (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**: 97–117.

Dalton, R. (2000). Biologists flock to 'evo-devo' in a quest to read the recipes of life. *Nature* **403**: 125.

Daly, H. V., J. T. Doyen, and A. H. Purcell III (1998). *Introduction to Insect Biology and Diversity.* 2nd Ed. Oxford Univ. Press, Oxford, UK.

Daniels, S. B., K. R. Peterson, L. D. Strausbaugh, M. G. Kidwell and A. Chovnick (1990). Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **124**: 339–355.

Davey, M. E., and G. A. O'Toole (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiol. Mol. Biol. Rev.* **64**: 847–867.

Davis, G. K., and N. H. Patel (1999). The origin and evolution of segmentation. *Trends Genet.* **15**: M68–72.

DeGroote, F., G. Pont, D. Micard, and G. Picard (1989). Extrachromosomal circular DNAs in *Drosophila melanogaster*: comparison between embryos and KcO% cells. *Chromosoma* **98**: 201–206.

Delpuech, J. M., C. F. Aquadro, and R. T. Roush (1993). Noninvolvement of the long terminal repeat of transposable element *17.6* in insecticide resistance in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **90**: 5643–5647.

De Robertis, E. M., and Y. Sasai (1996). A common plan for dorsoventral patterning in Bilateria. *Nature* **380**: 37–40.

DeVault, J. D., and S. K. Narang (1994). Transposable elements in Lepidoptera: hobo-like transposons in *Heliothis virescens* and *Helicoverpa zea*. *Biochem. Biophys. Res. Commun.* **203**: 169–175.

Devonshire, A. L., and L. M. Field (1991). Gene amplification and insecticide resistance. *Annu. Rev. Entomol.* **36**: 1–23.

Dobson, S. L., K. Bourtzis, H. R. Braig, B. F. Jones, W. Zhou, F. Rousset, and S. C. O'Neill (1999). *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* **29**: 153–160.

Douglas, A. E. (1992). Symbiotic microorganisms in insects. *Encycl. Microbiol.* **4**: 165–178.

Douglas, A. E. (1998). Nutritional interactions in insect–microbial symbioses: aphids and their symbiotic bacteria. *Annu. Rev. Entomol.* **43**: 17–37.

Durajlija Zinic, S., D. Ugarkovic, L. Cornudella, and M. Plohl (2000). A novel interspersed type of organization of satellite DNAs in *Tribolium madens* heterochromatin. *Chromosome Res.* **8**: 201–212.

Edgar, B. A., and T. L. Orr-Weaver (2001). Endoreplication cell cycles: more for less. *Cell* **105**: 297–306.

Eickbush, T. H., and W. D. Burke (1985). Silkmoth chorion gene families contain patchwork patterns of sequence homology. *Proc. Natl. Acad. Sci. USA* **82**: 2814–2818.

Eickbush, D. G., T. H. Eickbush, and J. H. Werren (1992). Molecular characterization of repetitive DNA sequences from a B chromosome. *Chromosoma* **101**: 575–583.

Erwin, D., J. Valentine, and D. Jablonski (1997). The origin of animal body plans. *Am. Sci.* **85**: 126–137.

Fallon, A. M. (1984). Methotrexate resistance in cultured mosquito cells. *Insect Biochem.* **14**: 697–704.

Feder, M. E., and R. A. A. Krebs (1997). Ecological and evolutionary physiology of heat shock proteins and the stress response in *Drosophila*: complementary insights from genetic engineering and natural variation, pp. 155–173. In: *Environmental Stress, Adaptation, and Evolution*. R. Bijlsma, and V. Loeschke, Eds. Birkhauser Verlag, Basel, Switzerland.

Field, L. M. (2000). Methylation and expression of amplified esterase genes in the aphid *Myzus persicae* (Sulzer). *Biochem. J.* **348**: 863–868.

Finnegan, D. J. (1990). Transposable elements, pp. 371–382. In: *The Genome of Drosophila melanogaster. Part 4: Genes L-Z, Balancers, Transposable elements*. Drosophila Inform. Service. Vol. (68).

Fitch, D. H. A., L. D. Strausbaugh, and V. Barrett (1990). On the origins of tandemly repeated genes: Does histone gene copy number in *Drosophila* reflect chromosomal location? *Chromosoma* **99**: 118–124.

Fleming, J. G. W. (1992). Polydnaviruses: mutualists and pathogens. *Annu. Rev. Entomol.* **37**: 401–425.

Fleming, J. G. W., and M. D. Summers (1991). Polydnavirus DNA is integrated in the DNA of its parasitoid wasp host. *Proc. Natl. Acad. Sci. USA* **88**: 9770–9774.

Frank, J. (2000). The ribosome—a macromolecular machine par excellence. *Chem. Biol.* **7**: R133–141.

Fraser, M. J. (1985). Transposon-mediated mutagenesis of baculoviruses: transposon shuttling and implications for speciation. *Ann. Entomol. Soc. Am.* **79**: 773–783.

Furlong, E. E.M., E. C. Andersen, B. Null, K. P. White, and M. P. Scott (2001). Patterns of gene expression during *Drosophila* mesoderm development. *Science* **293**: 1629–1633.

Garner, K. J., and J. M. Slavicek (1999). Identification of a non-LTR retrotransposon from the gypsy moth. *Insect Mol. Biol.* **8**: 231–242.

Gaubatz, J. W. (1990). Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryotic cells. *Mut. Res.* **237**: 271–292.

Gehring, W. J. (1985). The homeo box: a key to the understanding of development? *Cell* **40**: 3–5.

Gehring, W. J. (1987). Homeo boxes in the study of development. *Science* **236**: 1245–1252.

Gillespie, J. P., M. R. Kanost, and T. Trenczek (1997). Biological mediators of insect immunity. *Annu. Rev. Entomol.* **42**: 611–643.

Gray, M. W. (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**: 25–50.

Haig, D. (1993). The evolution of unusual chromosomal systems in sciarid flies: intragenomic conflict and the sex ratio. *J. Evol. Biol.* **6**: 249–261.

Hales, D. F., J. Tomiuk, K. Wohrmann, and P. Sunnucks (1997). Evolutionary and genetic aspects of aphid biology: a review. *Eur. J. Entomol.* **94**: 1–55.

Hall, B. K. (2000). Evo-devo or devo-evo—does it matter? *Evol. Dev.* **2**: 177–178.
Handler, A. M., and S. P. Gomez (1997). A new *hobo*, *Ac*, *Tam2* transposable element, *hopper*, from *Bactrocera dorsalis* is distantly related to *hobo* and *Ac Gene* **185**: 133–135.
Hankeln, T., and E. R. Schmidt (1990). New foldback transposable element TFB1 found in histone genes of the midge *Chironomus thummi*. *J. Mol. Biol.* **215**: 477–482.
Hankeln, T., and E. R. Schmidt (1991). The organization, localization, and nucleotide sequence of the histone genes of the midge *Chironomus thummi*. *Chromosoma* **101**: 25–31.
Hartl, D. L., and S. W. Brown (1970). The origin of male haploid genetic systems and their expected sex ratio. *Theoret. Pop. Biol.* **1**: 165–190.
Hawley, R. S., and M. Y. Walker (2000). *Drosophila melanogaster*: A genome sequenced. *Genome Res.* **10**: 593–594.
Heath, B. D., R. D. J. Butcher, W. G. F. Whitfield, and S. F. Hubbard (1999). Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* **9**: 313–316.
Heck, M. M. S., and A. C. Spradling (1990). Multiple replication origins are used during *Drosophila* chorion gene amplification. *J. Cell Biol.* **110**: 903–914.
Heckel, D. G. (1993). Comparative genetic linkage mapping in insects. *Annu. Rev. Entomol.* **38**: 381–408.
Heddi, A., A. M. Grenier, C. Khatchadourian, H. Charles, and P. Nardon (1999). Four intracellular genomes direct weevil biology: nuclear mitochondrial, principal endosymbiont, and *Wolbachia. Proc. Natl. Acad. Sci. USA* **95**: 6814–6819.
Heddi, A., H. Charles, and C. Khatchadourian (2001). Intracellular bacterial symbiosis in the genus *Sitophilus*: the 'biological individual' concept revisited. *Res. Microbiol.* **152**: 431–437.
Hill, S. R., S. S. Leung, N. L. Quercia, D. Vasiliauskas, J. Yu, I. Pasic, D. Leung, A. Tran, and P. Romans (2001). *Ikirara* insertions reveal five new *Anopheles gambiae* transposable elements in islands of repetitious sequence. *J. Mol. Evol.* **52**: 215–231.
Holland, P. W. H. (1999). The future of evolutionary developmental biology. *Nature* **402**: C41–44.
Houck, M. A., J. B. Clark, K. R. Peterson, and M. G. Kidwell (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **253**: 1125–1129.
Hoy, M. A. (2000). Deploying transgenic arthropods in pest management programs: Risks and realities, pp. 335–367. In: *Insect Transgenesis. Methods and Applications*. A. M. Handler and A. A. James, Eds. CRC Press, Boca Raton, FL.
Huijser, P., C. J. Kirchhoff, D. H. Lankenau, and W. Hennig (1988). Retrotransposon-like sequences are expressed in the Y chromosomal lampbrush loops of *Drosophila hydei*. *J. Mol. Biol.* **203**: 689–697.
Hultmark, D. (1993). Immune reactions in *Drosophila* and other insects: a model for innate immunity. *Trends Genet.* **9**: 178–183.
Hunt, G. J., and R. E. Page (1995). Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics* **139**: 1371–1382.
Hurst, G. D. D., and M. Schilthuizen (1998). Selfish genetic elements and speciation. *Heredity* **80**: 2–8.
Ichimura, S., K. Mita, and K. Sugaya (1997). A major non-LTR retrotransposon of *Bombyx mori*, L1Bm. *J. Mol. Evol.* **45**: 253–264.
Jabbari, K., and G. Bernardi (2000). The distribution of genes in the *Drosophila* genome. *Gene* **247**: 287–292.
Jakubczak, J. L., W. D. Burke, and T. H. Eickbush (1991). Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc. Natl. Acad. Sci. USA* **88**: 3295–3299.
Janies, D., and R. DeSalle (1999). Development, evolution, and corroboration. *Anatom. Rec. (New Anat.)* **257**: 6–14.
Jenner, R. A. (2000). Evolution of animal body plans: the role of metazoan phylogeny at the interface between pattern and process. *Evol. Dev.* **2**: 208–221.
Jeyaprakash, A., and M. A. Hoy (1995). Complete sequence of a *mariner* transposable element from the predatory mite *Metaseiulus occidentalis* isolated by an inverse PCR approach. *Insect Mol. Biol.* **4**: 31–39.
Jeyaprakash, A., and M. A. Hoy (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of 63 arthropod species. *Insect Mol. Biol.* **9**: 393–405.
Kafatos, F. C. (1981). Structure, evolution, and developmental expression of the silkmoth chorion multigene families. *Am. Zool.* **21**: 707–714.
Kafatos, F. C., C. Delidakis, W. Orr, G. Thireos, K. Komitopoulou and Y. Wong (1986). Studies on the developmentally regulated amplification and expression of *Drosophila* chorion genes, pp. 85–101. In: *Molecular Developmental Biology*, Alan R. Liss, New York.
Kafatos, F. C., C. Louis, C. Savakis, D. M. Glover, M. Ashburner, A. J. Link, I. Siden-Kiamos, and R. D. C. Saunders (1991). Integrated maps of the *Drosophila* genome: progress and prospects. *Trends Genet.* **7**: 155–161.

Kapitonov, V. V., and J. Jurka (2001). Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Sci. USA* **98**: 8714–8719.

Kaufman, M. G., E. D. Walker, D. A. Odelson, and M. J. Klug (2000). Microbial community ecology and insect nutrition. *Am. Entomol.* **46**: 173–184.

Kaulenas, M. S. (1985). Molecular Biology: protein synthesis, pp. 255–305. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology*. Vol. (10). G. A. Kerkut, and L. I. Gilbert, Eds. Pergamon, Oxford, UK.

Kerrebrock, A. W., R. Srivastava, and S. A. Gerbi (1989). Isolation and characterization of ribosomal DNA variants from *Sciara coprophila*. *J. Mol. Biol.* **210**: 1–13.

Khush, R. S., and B. Lemaitre (2000). Genes that fight infection. What the *Drosophila* genome says about animal immunity. Trends Genet. **16**: 442–468.

Kidwell, M. G. (1992). Horizontal transfer. *Curr. Opin. Genet. Dev.* **2**: 868–873.

Kimura, K., T. Okumura, O. Ninaki, M. G. Kidwell, and K. Suzuki. 1993. Transposable elements in commercially useful insects: I. Southern hybridization study of silkworms and honeybees using *Drosophila* probes. *Jpn. J. Genet.* **68**: 63–71.

King, R. E., Ed. (1975). *Handbook of Genetics.* Vol.(3). *Invertebrates of Genetic Interest*. Plenum, New York.

Klingler, M. (1994). Segmentation in insects: how singular is *Drosophila*? *BioEssays* **16**: 391–392.

Kobayashi, S., R. Amikura, and M. Okada (1993). Presence of mitochondrial large ribosomal RNA outside mitochondria in germ plasm of *Drosophila melanogaster*. *Science* **260**: 1521–1524.

Komaki, K., and H. Ishikawa (1999). Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. *J. Mol. Evol.* **48**: 717–722.

Komaki, K., and H. Ishikawa (2000). Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. *Insect Biochem. Mol. Biol.* **30**: 253–258.

Konsolaki, M., K. Komitopoulou, P. P. Tolias, D. L. King, C. Swimmer, and F. C. Kafatos (1990). The chorion genes of the medfly, *Ceratitis capitata*, I; structural and regulatory conservation of the s36 gene relative to two *Drosophila* species. *Nucleic Acids Res.* **18**: 1731–1737.

Krell, P. J. (1991). The polydnaviruses: multipartite DNA viruses from parasitic Hymenoptera, pp. 141–177. In: *Viruses of Invertebrates*. E. Kurstak, Ed. Marcel Dekker, New York.

Langworthy, N. G., A. Renz, U. Mackenstedt, K. Henkle-Duhrsen, M. B. D. Bronsvoort, V. N. Tanya, M. J. Donnelly, and A. J. Trees (2000). Macrofilaricidal activity of tetracycline against the filarial nematode *Onchocerca ochengi*: elimination of *Wolbachia* precedes worm death and suggests a dependent relationship. *Proc. Roy. S. Lond. B* **267**: 1063–1069.

Larsen-Rapport, E. W. (1986). Imaginal disc determination: molecular and cellular correlates. *Annu. Rev. Entomol.* **31**: 145–175.

Lauzon, C. R., R. E. Sjogren, S. E. Wright, and R. J. Prokopy (1998). Attraction of *Rhagoletis pomonella* (Diptera: Tephritidae) flies to odor of bacteria: apparent confinement to specialized members of Enterobacteriaceae. *Environ. Entomol.* **27**: 853–857.

Lauzon, C. R., R. E. Sjogren, and R. J. Prokopy (2000). Enzymatic capabilities of bacteria associated with apple maggot flies: a postulated role in attraction. *J. Chem. Ecol.* **26**: 953–967.

Lawrence, P. A. (1992). *The Making of a Fly. The Genetics of Animal Design*. Blackwell Scientific, Oxford, UK.

Lewis, E. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **176**: 141–152.

Lewis, E. B. (1994). Homeosis: the first 100 years. *Trends Genet.* **10**: 341–343.

Lilburn, T. G., T. M. Schmidt, and J. A. Breznak (1999). Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* **1**: 331–345.

Lu, L., H. Zhang, and J. Tower (2001). Functionally distinct, sequence-specific replicator and origin elements are required for *Drosophila* chorion gene amplification. *Genes Dev.* **15**: 134–146.

Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**: 595–605.

Mannervik. M. (1999). Target genes of homeodomain proteins. *BioEssays* **21**: 267–270.

Martin, W. (1999). A briefly argued case that mitochondria and plastids are descendants of endosymbionts, but that the nuclear compartment is not. *Proc. R. Soc.* London B **266**: 1387–1395.

Mason, J. M., A. Haoudi, A. Y. Konev, E. Kurenova, M. F. Walter, and H. Biessmann (2000). Control of telomere elongation and telomeric silencing in *Drosophila melanogaster*. *Genetica* **109**: 61–70.

Masui, S., S. Kamoda, T. Sasaki, and H. Ishikawa (1999). The first detection of the insertion sequence ISW1 in the intracellular reproductive parasite *Wolbachia*. *Plasmid* **42**: 13–19.

Masui, S., S. Kamoda, T. Sasaki, and H. Ishikawa (2000). Distribution and evolution of bacteriophage WO in *Wolbachia*, the endosymbiont causing sexual alterations in arthropods. *J. Mol. Evol.* **51**: 491–497.

Mayer, M. P., and B. Bukau (1999). Molecular chaperones: the busy life of *Hsp90*. *Curr. Biol.* **9**: R322–325.
Mikitani, K., T. Sugasaki, T. Shimada, M. Kobayashi, and J. A. Gufstafsson (2000). The chitinase gene of the silkworm, *Bombyx mori*, contains a novel *Tc*-like transposable element. *J. Biol. Chem.* **275**: 37725–37732.
Miller, W. J., J. F. McDonald, and W. Pinsker (1997). Molecular domestication of mobile elements. *Genetica* **100**: 261–270.
Mitsialis, S. A., and F. C. Kafatos (1985). Regulatory elements controlling chorion gene expression are conserved between flies and moths. *Nature* **317**: 453–456.
Mitsialis, S. A., S. Veletza, and F. C. Kafatos (1989). Transgenic regulation of moth chorion gene promoters in *Drosophila*: tissue, temporal, and quantitative control of four bidirectional promoters. *J. Mol. Evol.* **29**: 486–495.
Mizrokhi, L. J. and Mazo, A. M. (1990). Evidence for horizontal transmission of the mobile element *jockey* between distant *Drosophila* species. *Proc. Natl. Acad. Sci. USA*. **87**: 9216–9220.
Moran, N., A., and P. Baumann (2000). Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* **3**: 270–275.
Morata, G. (2001). How *Drosophila* appendages develop. *Nat. Rev. Mol. Cell Biol.* **2**: 89–98.
Morimoto, R. I., K. D. Sarge, and K. Abravaya (1992). Transcriptional regulation of heat shock genes. A paradigm for inducible genomic responses. *J. Biol. Chem.* **31**: 21987–21990.
Mouches, C., Y. Pauplin, M. Agarwal, L. Lemieux, M. Herzog, M. Abadon, V. Beyssat-Arnaouty, O. Hyrien, B. R. deSaint Vincent, G. P. Georghiou, and N. Pasteur (1990). Characterization of amplification core and esterase B1 gene responsible for insecticide resistance in *Culex*. *Proc. Natl. Acad. Sci. USA* **87**: 2574–2578.
Mouches, C., M. Agarwal, K. Campbell, L. Lemieux, and M. Abadon (1991). Sequence of a truncated LINE-like retroposon dispersed in the genome of *Culex* mosquitoes. *Gene* **106**: 279–280.
Mouches, C., N. Bensaadi, and J. C. Salvado (1992). Characterization of a LINE retroposon dispersed in the genome of three non-sibling *Aedes* mosquito species. *Gene* **120**: 183–190.
Mousseau, T. A., and H. Dingle (1991). Maternal effects in insect life histories. *Annu. Rev. Entomol.* **36**: 511–534.
Nur, U. (1990). Heterochromatization and euchromatization of whole genomes in scale insects (Coccoidea: Homoptera). *Development* 1990 Suppl.: 29–34.
Nusslein-Volhard, C., and E. Wieschaus (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* **187**: 795–801.
O'Brochta, D. A., W. D. Warren, K. J. Saville, and P. W. Atkinson (1996). *Hermes*, a functional non-drosophilid gene vector from *Musca domestica*. *Genetics* **142**: 907–914.
Ohno, S., and T. Yomo (1991). The grammatical rule for all DNA: Junk and coding sequences. *Electrophoresis* **12**: 103–108.
Okazaki, S., K. Tsuchida, H. Maekawa, H. Ishikawa, and H. Fujiwara (1993). Identification of a pentanucleotide telomeric sequence, $(TTAGG)_n$, in the silkworm *Bombyx mori* and in other insects. *Mol. Cell. Biol.* **13**: 1424–1432.
Okazaki, S., H. Ishikawa, and H. Fujiwara (1995). Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm *Bombyx mori*. *Mol. Cell. Biol.* **15**: 4545–4552.
O'Neill, S. L., R. Giordano, A. M. E. Colbert, T. L. Karr, and H. M. Robertson (1992). 16S rRNA phylogenetic analysis of the bacterial symbiont associated with cytoplasmic incompatibility in insects. *Proc. Natl. Acad. Sci. USA* **89**: 2699–2702.
O'Neill, S. L., A. A. Hoffmann, and J. H. Werren, Eds. (1997). *Influential Passengers. Inherited Microorganisms and Arthropod Reproduction*, Oxford Univ. Press, UK.
Orr-Weaver, T. L. (1991). *Drosophila* chorion genes: cracking the eggshell's secrets. *BioEssays* **113**: 97–105.
Otto, P. A. (2000). *Drosophila* viewer: a program on the formal genetics, anatomy and developmental biology of *Drosophila melanogaster* for students and specialists. *Gen. Mol. Biol.* **23**: 835–839.
Otto, S. P., and J. Whitton. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**: 401–437.
Pardue, M. L. (1991). Dynamic instability of chromosomes and genomes. *Cell* **66**: 427–431.
Pardue, M. L., and W. Hennig (1990). Heterochromatin: junk or collectors item? *Chromosoma* **100**: 3–7.
Pasteur, N., and M. Raymond (1996). Insecticide resistance genes in mosquitoes: their mutations, migration, and selection in field populations. *J. Hered.* **87**: 444–449.
Paton, M. G., S. H. P. P. Karunaratne, E. Giakoumaki, N. Roberts, and J. Hemingway (2000). Quantitative analysis of gene amplification in insecticide-resistant *Culex* mosquitoes. *Biochem. J.* **346**: 17–24.
Pauli, D., A. P. Arrigo, and A. Tissieres (1992). Heat shock response in *Drosophila. Experientia* **48**: 623–629.
Pennisi, E., and W. Roush (1997). Developing a new view of evolution. *Science* **277**: 34–37.
Perkins, H. D., and A. J. Howells (1992). Genomic sequences with homology to the P element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina*. *Proc. Natl. Acad. Sci. USA* **89**: 10753–10757.

Perrimon, N., D. Mohler, L. Engstrom, and A. Mahowald (1986). X-linked female-sterile loci in *Drosophila melanogaster*. *Genetics* **113**: 695–712.

Pfeifer, T. A., M. Ring, and T. A. Grigliatti (2000). Identification and analysis of *Lydia*, a LTR-retrotransposon from *Lymantria dispar*. *Insect Mol. Biol.* **9**: 349–356.

Pinkerton, A. C., S. Whyard H. A. Mende, C. J. Coates, D. A. O'Brochta, and P. W. Atkinson (1999). The Queensland fruit fly, *Bactrocera tryoni*, contains multiple members of the *hAT* family of transposable elements. *Insect Mol. Biol.* **8**: 423–434.

Plasterk, R. H. A., Z. Izsvak, and Z. Ivics (1999). Resident aliens. The *Tc1/mariner* superfamily of transposable elements. *Trends Genet.* **15**: 326–332.

Powell, J. R. (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford Univ. Press, New York.

Prody, C. A., P. Dreyfus, R. Zamir, H. Zakut, and H. Soreq (1989). *De novo* amplification within a 'silent' human cholinesterase gene in a family subjected to prolonged exposure to organophosphorous insecticides. *Proc. Natl. Acad. Sci. USA* **86**: 690–694.

Prudhomme, J. C., P. Couble, J. P. Garel, and J. Daillie (1985). Silk synthesis, pp. 571–594. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology.* Vol. (10). G. A. Kerkut and L. I. Gilbert, Eds. Pergamon, Oxford, UK.

Rai, K. S., and W. C. Black, IV (1999). Mosquito genomes: structure, organization and evolution. *Adv. Genet.* **41**: 1–33.

Raikhel, A. S., and T. S. Dhadialla (1992). Accumulation of yolk proteins in insect oocytes. *Annu. Rev. Entomol.* **37**: 217-251.

Retnakaran, A., and J. Percy (1985). Fertilization and special modes of reproduction, pp. 231–293. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology.* Vol. (1). G. A. Kerkut and L. I. Gilbert, Eds. Pergamon, Oxford, UK.

Rigaud, T. (1999). Further *Wolbachia* endosymbiont diversity: a tree hiding in the forest? *Trends Ecol. Evol.* **14**: 212–213.

Rigaud, T., and F. Rousset (1996). What generates the diversity of *Wolbachia*–arthropod interactions? *Biodiversity Conserv.* **5**: 999–1013.

Rina, M., and C. Savakis (1991). A cluster of vitellogenin genes in the Mediterranean fruit fly *Ceratitis capitata*: sequence and structural conservation in dipteran yolk proteins and their genes. *Genetics* **127**: 769–780.

Rinderer, T. E., Ed. (1986). *Bee Genetics and Breeding*. Academic Press, New York.

Robertson, H. M. (1993). The *mariner* transposable element is widespread in insects. *Nature* **362**: 241–245.

Robertson, H. M., and D. J. Lampe (1995). Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **40**: 333–357.

Robertson, H. M., D. J. Lampe, and E. G. MacLeod (1992). A *mariner* transposable element from a lacewing. Nucleic Acids Res. **20**: 6409.

Robinson, R. (1971). *Lepidoptera Genetics*. Pergamon, Oxford, UK.

Rokas, A. (2000). *Wolbachia* as a speciation agent. *Trends Ecol. Evol.* **15**: 44–45.

Rutherford, S. L., and S. Lindquist (1998). Hsp90 as a capacitor for morphological evolution. *Nature* **396**: 336–342.

Sapienza, C., A. C. Peterson, J. Rossant, and R. Balling (1987). Degree of methylation of transgenes is dependent on gamete of origin. *Nature* **328**: 251–254.

Schwemmler, W., and G. Gassner, Eds. (1989). *Insect Endocytobiosis: Morphology, Physiology, Genetics, Evolution*. CRC Press, Boca Raton, FL.

Severson, D. W., A. Mori, Y. Zhang, and B. M. Christensen (1993). Linkage map for *Aedes aegypti* using restriction fragment length polymorphisms. *J. Hered.* **84**: 241–247.

Severson, D. W., S. E. Brown, and D L. Knudson (2001). Genetic and physical mapping in mosquitoes: molecular approaches. *Annu. Rev. Entomol.* **46**: 183–219.

Sezutsu, H., and K. Yukuhiro (2000). Dynamic rearrangement within the *Antheraea pernyi* silk fibroin gene is associated with four types of repetitive units. *J. Mol. Evol.* **51**: 329–338.

Shapiro, J. A. (1999). Transposable elements as the key to a 21st century view of evolution. *Genetica* **107**: 171–179.

Shelby, K. S., and B. A. Webb (1999). Polydnavirus-mediated suppression of insect immunity. *J. Insect Physiol.* **45**: 507–514.

Shi, J., D. G. Heckel, and M. R. Goldsmith (1995). A genetic linkage map for the domesticated silkworm, *Bombyx mori*, based on restriction fragment length polymorphisms. *Genet. Res.* **66**: 109–126.

Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**: 81–86.

Shoemaker, D.D., V. Katju, and J. Jaenike (1999). *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria*. *Evolution* **53**: 1157–1164.

Snook, R. R., S. Y. Cleland, M. F. Wolfner, and T. L. Karr (2000). Offsetting effects of *Wolbachia* infection and heat shock on sperm production in *Drosophila simulans*: analyses of fecundity, fertility, and accessory gland proteins. *Genetics* **155**: 167–178.

Sokoloff, A. (1966). *The Genetics of* Tribolium *and Related Species.* Academic Press, New York.

Sokoloff, A. (1977). *The Biology of* Tribolium*, with Special Emphasis on Genetic Aspects.* Vol. (3). Clarendon, Oxford, UK.

Solter, D. (1988). Differential imprinting and expression of maternal and paternal genomes. *Annu. Rev. Genet.* **22**: 127–146.

Staiber, W. (1987). Unusual germ line limited chromosomes in *Acricotopus lucidus* (Diptera, Chironomidae). *Genome* **29**: 702–705.

Stark, G. R., and G. M. Wahl (1984). Gene amplification. *Annu. Rev. Biochem.* **53**: 447–491.

Steiner, W. W. M., W. J. Tabachnick, K. S. Rai, and S. Narang, Eds. (1982). *Recent Developments in the Genetics of Insect Disease Vectors*. Stipes Publ., Champaign, IL.

Stoltz, D., and J. B. Whitfield (1992). Viruses and virus-like entities in the parasitic Hymenoptera. *J. Hymen. Res.* **1**: 125–139.

Suck, G., and W. Traut (2000). TROMB, a new retrotransposon of the *gypsy-Ty3* group from the fly *Megaselia scalaris*. *Gene* **255**: 51–57.

Sun, L. V., J. M. Foster, G. Tzertzinis, M. Ono, C. Bandi. B. E. Slatko, and S. L. O'Neill (2001). Determination of *Wolbachia* genome size by pulse-field gel electrophoresis. *J. Bacteriol.* **183**: 2219–2225.

Takahashi, H., and H. Fujiwara (1999). Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyx mori*. *Nucleic Acids Res.* **27**: 2015–2021.

Tan, Y.D., C. Wan, Y. Zhu, C. Lu, Z. Xiang, and H. W. Deng (2001). An amplified fragment length polymorphism map of the silkworm. *Genetics* **157**: 1277–1284.

Tazima, Y. (1964). *Genetics of the Silkworm*. Logos, London, UK.

Tazima, Y., H. Doira, and H. Akai (1975). The domesticated silkmoth, *Bombyx mori*, pp. 63–124. In: *Handbook of Genetics.* Vol. 3. *Invertebrates of Genetic Interest*. R.C. King, Ed. Plenum, New York.

Thomas, B. J., and D. A. Wassarman (1999). A fly's eye view of biology. *Trends Genet.* **15**: 184–190.

Trewitt, P. M., L. J. Heilmann, S. S. Degrugillier, and A. K. Kumaran (1992). The boll weevil vitellogenin gene: nucleotide sequence, structure, and evolutionary relationship to nematode and vertebrate vitellogenin genes. *J. Mol. Evol.* **34**: 478–492.

Tu, Z. (2000). Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **98**: 1699–1704.

Tu, Z., J. Isoe, and J. A. Guzova (1998). Structural, genomic and phylogenetic analysis of *Lian*, a novel family of non-LTR retrotransposons in the yellow fever mosquito, *Aedes aegypti*. *Mol. Biol. Evol.* **15**: 837–853.

van Meer, M. M. M., and R. Stouthamer (1999). Cross-order transfer of *Wolbachia* from *Muscidifurax uniraptor* (Hymenoptera: Pteromalidae) to *Drosophila simulans* (Diptera: Drosophilidae). *Heredity* **82**: 163–169.

von Dassow, G., and E. Munro (1999). Modularity in animal development and evolution: elements of a conceptual framework for EvoDevo. *J. Exp. Zool. (Mol. Dev. Evol.)* **285**: 307–325.

von Dohlen, D. C., S. Kohler, S. T. Alsop, and W. R. McManus (2001). Mealybug B-proteobacterial endosymbionts contain g-proteobacterial symbionts. *Nature* **412**: 433–435.

Wagner, R. P., M. P. Maguire, and R. L. Stallings (1993). *Chromosomes: A Synthesis*. Wiley-Liss, New York.

Warren, A. M., and J. M. Crampton (1991). The *Aedes aegypti* genome: complexity and organization. *Genet. Res. Camb.* **58**: 225–232.

Warren, W. D., P. W. Atkinson, and D. A. O'Brochta (1994). The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genet. Res. Camb.* **64**: 87–97.

Waters, L. C., A. C. Zelhof, B. J. Shaw, and L. Y. Chang (1992). Possible involvement of the long terminal repeat of transposable element *17.6* in regulating expression of an insecticide resistance-associated P450 gene in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **89**: 4855–4859.

Webb, B. A. (1998). Polydnavirus biology, genome structure and evolution, pp. 105–139. In: *The Insect Viruses*. L. K. Miller and A. Ball, Eds. Plenum, New York.

Webb, B. A., and L. Cui (1998). Relationships between polydnavirus genomes and viral gene expression. *J. Insect Physiol.* **44**: 785–793.

Welburn, S. C., K. Arnold, I. Maudlin, and G. W. Gooday (1993). Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse flies. *Parasitology* **107**: 141–145.

Werren, J. H., L. R. Guo, and D. W. Windsor (1995). Distribution of *Wolbachia* in neotropical arthropods. *Proc. R. Soc. London* B **262**: 197–204.

West-Eberhard, M. J. (1998). Evolution in the light of developmental and cell biology, and *vice versa*. *Proc. Natl. Acad. Sci. USA* **95**: 8417–8419.

White, K. P., S. A. Rifkin, P. Hurban, and D. S. Hogness (1999). Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**: 2179–2184.

White, M. J. D. (1973). *Animal Cytology and Evolution*. 3rd Ed. Cambridge Univ. Press, Cambridge, UK.

Whitfield, J. B. (1997). Molecular and morphological data suggest a single origin of the polydnaviruses among braconid wasps. *Naturwissenschaften* **84**: 502–507.

Wilkins, A. S. (1993). *Genetic Analysis of Animal Development*. 2nd Ed. Wiley-Liss, New York.

Williams, S. M., and L. G. Robbins (1992). Molecular genetic analysis of *Drosophila* rDNA arrays. *Trends Genet.* **8**: 335–340.

Wilson, T. G. (1993). Transposable elements as initiators of insecticide resistance. *J. Econ. Entomol.* **86**: 645–651.

Wolf, K. W., H. G. Mertl, and W. Traut (1991). Structure, mitotic, and meiotic behaviour, and stability of centromere-like elements devoid of chromosome arms in the fly *Megaselia scalaris* (Phoridae). *Chromosoma* **101**: 99–108.

Wright, J. W., and R. Pal, Eds. (1967). *Genetics of Insect Vectors of Disease*. Elsevier, Amsterdam, The Netherlands.

Wu, C., S. Asakawa, N. Shimizu, S. Kawasaki, and Y. Yasukochi (1999). Construction and characterization of bacterial artificial chromosome libraries from the silkworm, *Bombyx mori*. *Mol. Gen. Genet.* **261**: 698–706.

Xiong, Y., W. D. Burke, and T. H. Eickbush (1993). *Pao*, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R region. *Nucleic Acids Res.* **21**: 2117–2123.

Yasukochi, Y. (1998). A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. *Genetics* **150**: 1513–1525.

Yoshida, N. K. Oeda, E. Watanabe, T. Mikami, Y. Fukita, K. Nishimura, K. Komai, and K. Matsuda (2001). Chaperonin turned insect toxin. *Nature* **411**: 44.

Zheng, L., R. D. C. Saunders, D. Fortini, A. della Torre, M. Coluzzi, D. M. Glover, and F. C. Kafatos (1991). Low-resolution genome map of the malaria mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **88**: 11187–11191.

Zheng, L., F. H. Collins, V. Kumar, and F. C. Kafatos (1993). A detailed genetic map for the X chromosome of the malaria vector, *Anopheles gambiae*. *Science* **261**: 605–608.

Zheng, L., M. Q. Benedict, A. J. Cornel, F. H. Collins, and F. C. Kafatos (1996). An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics* **143**: 941–952.

Zhou, W., F. Rousset, and S. L. O'Neill (1998). Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. R. Soc. London* B **265**: 509–515.

Some Relevant Web Sites

The Interactive Fly (http://sdb.bio.purdue.edu/fly/aimain/links): This site has links that provide information on development and genomes of *Drosophila*, *Drosophila* research laboratories, sites on gene families, journals, and all principal sources of information on *Drosophila*.

FlyBase. A Database of the *Drosophila* Genome (http://flybase.bio.indiana.edu/): This site includes cytological maps, annotated genome maps, lists of genes and gene products, gene expression data, Genome Projects' homepages, stock centers, information on transgene constructs, transposons found in *Drosophila*, literature, and people working on *Drosophila*.

www.fruitfly.org
www.celera.com
www.hgsc.bcm.tmc.edu/drosophila

5

Some Basic Tools: How to Cut, Paste, Copy, Measure, and Visualize DNA

5.1. Overview

Genetic engineers use a number of techniques to isolate DNA, cut and join molecules, and monitor the results. In order to clone a gene, determine its sequence, or alter the genetic makeup of an arthropod, various microbiological techniques are employed. The bacterium *Escherichia coli* has become a molecular biology workhorse because it can be induced to produce large amounts of recombinant DNA molecules by inserting plasmids, the bacteriophage λ, or genetically engineered variants of these agents into it.

Genetic engineers use a number of enzymes from different organisms to modify, ligate, or splice DNA. Purifying plasmids from *E. coli*, visualizing DNA by electrophoresis through agarose or polyacrylamide gels, conducting Southern blot analyses, and producing labeled probes are techniques that are basic to the molecular geneticist. As a mechanism to introduce these techniques, this chapter describes the steps involved in inserting a foreign gene into a plasmid, inserting the plasmid into *E. coli*, and isolating and analyzing the amplified DNA by Southern blot analysis and restriction site mapping. Northern and Western blot analyses allow the researcher to evaluate RNA and proteins, respectively. Many of these procedures have been simplified by the availability of commercial kits, but understanding the concepts behind them will allow you to “trouble shoot.”

5.2. Introduction to a Simple Experiment

A diverse array of molecular genetic techniques, some arising from research on apparently nonapplied topics, have become crucial tools that allow scientists to manipulate DNA from living organisms. This molecular genetic revolution began only about 30 years ago.

Prior to 1970, there was no way to cut a DNA molecule into discrete and predictable fragments, nor could specific DNA fragments be joined together. The discovery of enzymes called **restriction endonucleases** and **ligases** solved this. Much of genetic engineering technology is dependent upon our ability to cut DNA molecules at specific sites and combine them into new molecules by base pairing and ligation.

Another significant development was the harnessing of **plasmids** and **bacteriophages** as vehicles (vectors) to replicate foreign DNA (=clone) within the bacterium *Escherichia coli*. Cloning allowed nearly unlimited amounts of specific DNA to be produced for study and manipulation. Several techniques for monitoring the results of such manipulations were developed so that researchers could identify changes in DNA molecules as small as a single base modification.

A simple cloning project, as would have been conducted in 1985 when many of the current kits were unavailable, is outlined in Figure 5.1. The methods described provide the theoretical background to many of the current protocols used, although details are changed and may not be explained for proprietary reasons.

Understanding the theoretical bases of the techniques described in this chapter provides a background helpful in understanding many of the procedures one carries out with commercially available kits. Whenever using kits, carefully read the methods provided by the manufacturer. Examine each step to be sure you understand what you are doing and why before you start. If you have questions, most companies marketing kits provide a “hot line” that you can telephone for advice and information. Because improvements in methods are made regularly, it is a good idea to contact the manufacturer’s hot line to check on methods updates; some won’t be in the printed directions.

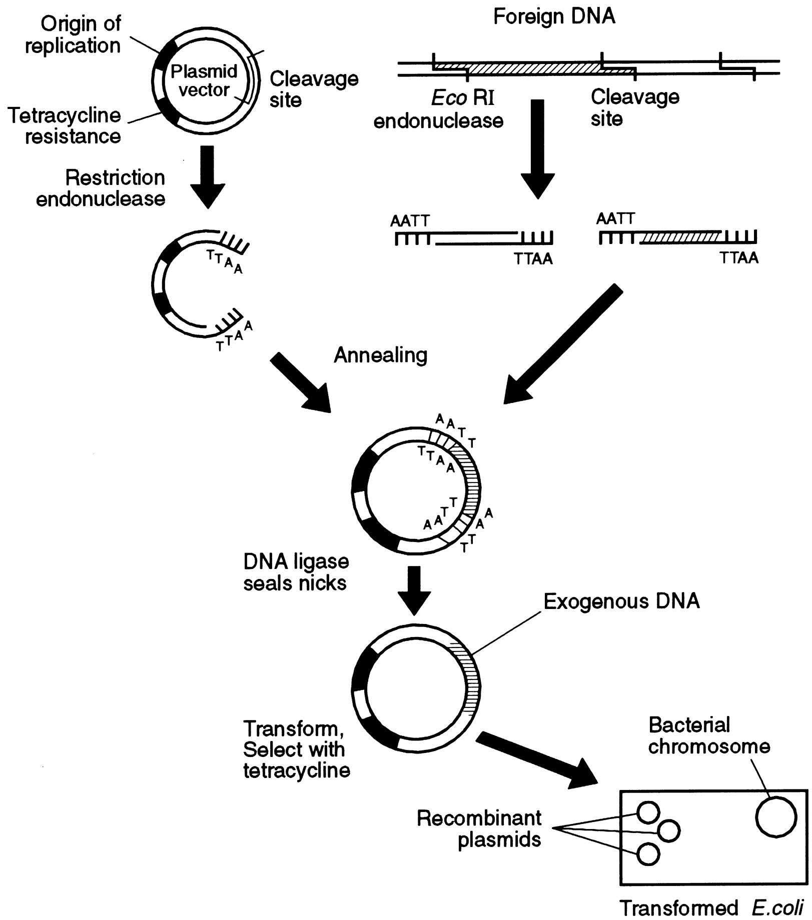


Figure 5.1. Outline of a simple cloning project involving insertion of foreign DNA into a plasmid vector, and its subsequent insertion into a bacterial host, *E. coli*, to produce multiple copies (clones) of the foreign DNA in recombinant plasmids. Both foreign and plasmid DNA is cut with a restriction endonuclease to produce ends that will allow annealing of the plasmid and foreign DNA fragments. The addition of DNA ligase combines the two DNA molecules, and the plasmid is inserted into *E. coli* where it will produce multiple copies of the new, recombinant DNA molecule. Subsequently, the plasmids will be extracted from their host cells and used for other studies.

The project described here is simple and basic (Figure 5.1). It involves inserting a piece of **exogenous** (or foreign) DNA extracted from one organism into a plasmid that has been engineered to serve as a **vector** to carry the exogenous DNA into *E. coli*. The *E. coli* cells with the exogenous DNA (in the plasmid vector) will be mass produced to yield large numbers of the desired DNA molecules. Once the DNA has been mass produced (**cloned**), the exogenous DNA can be studied in detail.

Although the experiment in Figure 5.1 conceptually is very simple, cloning a fragment of foreign DNA in a vector demands that several steps be achieved: 1) The circular vector DNA must be purified and cut. 2) The exogenous DNA must be extracted, purified, and cut. 3) The vector DNA and exogenous DNA must be joined together. 4) The reactions

should be monitored. 5) The recombinant plasmid or vector containing the exogenous DNA must be put back into *E. coli* to be amplified. 6) The recombinant plasmid then must be removed from *E. coli* and purified for analysis or use of the exogenous DNA.

This chapter provides an introduction to the procedures that could be employed to carry out the experiment in Figure 5.1. Simplified protocols of some procedures are provided for those interested in knowing something about the steps involved, although the methods described are illustrative rather than complete laboratory protocols. There are many excellent laboratory manuals available that provide detailed techniques (Sambrook and Russell 2001) as well as Web sites with up-to-the-minute protocols. As noted above, many of the techniques have been simplified and are available in kits provided by commercial sources. This has led one molecular biologist to note the availability of these kits has led to a "paradigm shift" in biology (Gilbert 1991).

5.3. Extracting DNA

The DNA to be manipulated must be extracted from an intact organism or from cells. This DNA also must be purified before it can be used. The degree of purity needed is determined by the goals of the experiment. Any time DNA is to be cut or ligated or cloned it must be very clean, meaning that molecules (such as proteins, lipids, sugars) that could interfere with these procedures have been removed.

One of the most common methods formerly used for extracting and purifying nucleic acids used phenol (Table 5.1). First cells are disrupted mechanically by grinding to release the genomic DNA. The insect sample usually is ground in a solution containing protease K (Jowett 1986, Sambrook and Russell 2001). Once the DNA is released, phenol can be used to extract DNA (or RNA) in large- or small-scale procedures. A plethora of different phenol extraction methods have been published, but the primary function of phenol is to remove proteins from an aqueous solution containing nucleic acids. Proteins need to be removed because some of them may be nucleases that could damage the DNA, while others simply could interfere with later manipulations. EDTA (ethylenediaminetetraacetic acid) is

Table 5.1. Rapid Phenol Extraction of Genomic DNA from *D. melanogaster*

1. Homogenize 50 to 200 flies (frozen in liquid nitrogen) in a 15 ml polypropylene tube with a Teflon pestle in 2 ml of lysis buffer.
   [Lysis buffer contains 100 mM Tris–HCl (pH 8.0), 50 mM NaCl, 50 mM EDTA, 1% SDS, 0.15 mM spermine, 0.5 mM spermidine.]
2. Add 20 μl of proteinase K solution (10 mg/ml).
3. Leave at 37°C for 1–2 h, but occasionally swirl and invert the tube to mix.
4. Extract once with an equal volume of phenol + chloroform + isoamyl alcohol.
   (The phenol, chloroform, and isoamyl alcohol should be in a 24:24:1 ratio. The isoamyl alcohol serves as an antifoaming agent.)
5. Spin in a bench centrifuge for 5 min at room temperature.
6. Decant the aqueous layer with a Pasteur pipette into a new tube.
7. Extract twice more with phenol + chloroform + isoamyl alcohol.
   Respin and decant the aqueous layer each time.
8. Extract the aqueous layer with chloroform and isoamyl alcohol (24:1).
   The interface between the organic and aqueous layer should now be clean.

Modified from Jowett (1986).

often added; it is a chelating agent that binds magnesium (Mg^{2+}) ions that are required for nucleases to act on the DNA.

Highly purified phenol is mixed with the sample under conditions that favor the dissociation of proteins from the nucleic acids, and the sample is then centrifuged (Table 5.1). Centrifugation yields two phases: 1) a lower organic phenol phase carrying the protein, and 2) the less dense aqueous phase containing the nucleic acids. Highly purified liquefied phenol ready for use in DNA extraction and purification can be purchased. Some phenol extraction protocols include chloroform, which denatures proteins, removes lipids, and improves the efficiency of the extractions. To reduce foaming caused by chloroform, isoamyl alcohol also is usually added. Handle phenol with great care and use only in a fume hood; phenol is toxic.

Extraction of DNA from cells or organisms should be carried out as quickly as possible in ice with refrigerated buffers to minimize the activity of any nucleases present in the cells that can degrade the DNA.

5.4. Precipitating Nucleic Acids

During cloning, it is often necessary to concentrate DNA samples or change the solvent in which the nucleic acid is dissolved. DNA isolated by phenol contains trace amounts of phenol, which could disrupt the activity of enzymes in subsequent manipulations if it were not purified further. Purification can be achieved by ethanol (EtOH) precipitation, isopropanol precipitation, or several other methods. The most versatile is probably EtOH precipitation because it can concentrate both DNA and RNA and purify DNA after phenol extractions.

Basically, DNA is precipitated by combining the DNA sample, a salt, and EtOH at −20°C or lower (Table 5.2). The precipitated salt of the nucleic acid is then sedimented by centrifugation, the EtOH supernatant is removed, and the nucleic acid pellet is resuspended

Table 5.2. Sample Protocol for Precipitating DNA with Ethanol

For recovery of DNA from a typical reaction (1 µg DNA in 20 µl):

1. To 20 µl aqueous DNA sample in a microcentrifuge tube, add 2 µl 3 M sodium acetate pH 5.5, and 40 µl EtOH.
2. Mix well by vortexing and immerse the tube in a −70°C bath composed of methanol plus dry ice for 15 min. The mixture should freeze or form a slurry.
3. Centrifuge the DNA precipitate in a benchtop microcentrifuge at maximum speed for 10 min in a cold room. A whitish pellet of DNA should appear at the bottom of the tube. In general, pellets of 10 µg are visible, while pellets of 2 µg will be invisible.
4. Remove the EtOH supernatant using a micropipette, taking care not to disturb the pellet or the area of the tube where the pellet should be located.
5. Add 100 µl 70% EtOH (chilled to −20°C) to the sample and vortex. This step removes any solute trapped in the precipitate.
6. Reprecipitate the DNA by centrifugation for 2 min and remove the supernatant as before.
7. Dry the pelleted DNA for 1 to 2 min in a vacuum desiccator taking care to release the vacuum gently so as not to dislodge the dried sample.
8. Resuspend the DNA in TE (pH 8) buffer (TE buffer contains 10 mM Tris-HCl at pH 8; 1 mM Na_2EDTA) or in sterile water.

Modified from Berger and Kimmel (1987).

in a buffer. Which salt or which buffer is used is determined by the nature of the sample and by the planned use for the nucleic acid. Once the DNA is purified, it can be stored at 4°C in TE (pH 8) buffer or sterile water. EDTA in the TE buffer helps to prevent degradation of the DNA during storage because it chelates heavy metal ions which are commonly required for DNase activity. For very long term storage (5 years or more), the DNA can be frozen at −80°C but should not be subjected to freeze–thaw cycles or it will be damaged (sheared).

The now-purified DNA can be cut, either by shearing or with a restriction endonuclease. Shearing produces random fragments, whereas restriction endonucleases can generate fragments of a desired size and with termini, or ends, appropriate to the annealing and ligation steps of the experiment outlined in Figure 5.1.

5.5. Shearing DNA

A variety of protocols are available to mechanically produce fragmented DNA. DNA in cells can be broken by shear forces in solution. Sonication with ultrasound can produce DNA fragments about 300 nucleotides long. High-speed stirring of cells in a blender at 1500 revolutions/min for 30 min will produce DNA molecules with a mean size of about 8 kb. Breakage occurs essentially at random with respect to DNA sequence, and the broken ends consist of short, single-stranded regions. These single-stranded termini must be modified before the DNA can be joined to a vector, so mechanically sheared DNA is rarely used in experiments. More often, DNA is cut in a specific manner to make it easier to manipulate it.

5.6. Cutting DNA with Restriction Endonucleases

Most cloning projects use restriction endonucleases to cut DNA. Restriction enzymes were discovered as an outcome of basic research aimed at understanding how bacteria control infections by **bacteriophages** (viruses that invade bacteria).

Most bacteria contain a variety of specific endonucleases that guard against invasion of foreign DNA (Frank 1994). These endonucleases make cuts in double-stranded DNA invading the cell unless the DNA has been modified in a specific manner by methylation. Thus, "foreign DNA" has an inappropriate methylation pattern, but the bacterium's own DNA is protected because its methylation pattern is recognized as appropriate. More than 3400 restriction enzymes (restriction endonucleases) have been identified. Endonucleases with at least 196 different sequence specificities are commercially available (Pingoud et al. 1993, Kessler and Manta 1990; also see the REBASE 2000 Web site, in the list following the references at the end of this chapter).

Restriction enzymes are identified by three-letter abbreviations for the parent organism (for example, *Hin* for *Haemophilus influenzae* or *Bam* for *Bacillus amyloliquefaciens*). An additional letter is added, if needed, to identify a specific strain or serotype (*Hin*d or *Bam*H). A Roman numeral is added to reflect the order of identification or characterization of the specific endonuclease (*Hin*dIII or *Bam*HI, Table 5.3).

Restriction endonucleases recognize specific sequences in DNA. They cleave the DNA in a precise manner, producing either blunt or staggered cuts (Table 5.3). Most endonucleases recognize sequences 4 to 6 nucleotides long, but some have 7- and 8-base recognition sites, and a very few recognize 12-base sequences. Restriction enzymes cleave DNA to form 5′-phosphate and 3′-hydroxyl termini on each strand (Table 5.3). Endonucleases that

Table 5.3. Target Sites for Selected Restriction Endonucleases

Organism from which endonuclease was isolated	Abbreviation	Recognition sequences (I indicates cleavage site)
Bacillus amyloliquefaciens H	*Bam*HI	G I GATCC
Bacillus globigii	*Bgl*II	A I GATCT
Escherichia coli RY13	*Eco*RI	G I AATTC
Haemophilus aegyptius	*Hae*III	GG I CC
Klebsiella pneumoniae	*Kpn*II	GGTAC I C
Nocaardia otitidis-caviarum	*Not*I	GC I GGCCGC
Providencia stuartii	*Pst*I	CTGCA I G
Serratia marcescens	*Sma*I	CCC I GGG

Recognition sequences are written from 5′ to 3′ with only one strand given. Modified from Kessler and Manta (1990).

Table 5.4. Three Types of Termini Are Created by Restriction Enzyme Cleavage of Double-Stranded DNA: A 5′ Overhang, Blunt Ends, or a 3′ Overhang[a]

***Eco*RI**	***Pvu*II**	***Kpn*I**
G I AATTC CTTAA I G	CAG I CTG GTC I GAC	GGTAC I C C I CATGG
$G^{3'\ 5'}AATTC$ $CTTAA^{5'\ 3'}G$	$CAG^{3'\ 5'}CTG$ $GTC^{5'\ 3'}GAC$	$GGTAC^{3'\ 5'}C$ $C^{5'\ 3'}CATGG$
5′ overhang	**Blunt end**	**3′ overhang**

[a]The I indicates where cleavage occurs when DNA is restricted, or cut, by three representative enzymes.

produce staggered breaks generate either 5′-phosphate extensions or 3′-hydroxyl extensions. Other endonucleases produce "blunt" breaks (Table 5.4).

It is often desirable to generate DNA fragments of a specific length, with a specific sequence, and with a particular type of end. This precision is possible with well-characterized DNA that has been sequenced. Such precision is not possible with uncharacterized DNA, except to predict whether the ends will be blunt, or with 5′ or 3′ overhangs. It is difficult to predict precisely the length of the DNA fragments that will be generated after digesting unknown DNA sequences with a particular restriction endonuclease, although we can predict the sequences at the ends of each fragment (Table 5.4).

Predictable fragment lengths would occur if all DNA sequences contained 50% guanine and cytosine (G+C) base pairs, and if all bases were distributed randomly in the DNA. Under these conditions, a 4-base sequence recognized by the restriction endonuclease would occur approximately every 256 bases (4^4), a 6-base sequence would occur approximately every 4 kb (4^6 or 4096 bases), and an 8-base sequence would occur approximately every 65 kb. However, many segments of DNA are not random in their G+C content. For example, highly repetitive DNA may have several nucleotides repeated millions of times, which obviously would bias restriction site frequencies significantly. The percentage of G+C in DNA from different sources may vary from 22 to 73%.

Many different restriction endonucleases are commercially available in either native or cloned form. Most manufacturers provide standardized buffers (high, medium, or low salt)

for optimizing the reaction conditions and protocols for carrying out the digestions. Among those available, some recognize identical sequences, although they may vary with respect to their sensitivity to methylation and cleavage rates. Restriction endonucleases can be degraded if not properly stored at –20°C and should be aliquoted into small amounts so they do not undergo multiple freeze–thaw cycles when used.

Choosing which endonuclease to use is determined by the goals of the project. Enzymes that produce small segments of a few hundred bases are useful for restriction mapping or for sequencing. Enzymes that produce fragments of 1 to 10 kb are useful for mapping large DNA regions and for cloning whole genes with their introns and control sequences. Generating even larger fragments (5 to 50 kb) is necessary for cloning into cosmid vectors or for genome walking (as discussed in Chapters 6 and 7).

Digestion reactions with restriction endonucleases contain the DNA substrate, the restriction endonuclease(s), Tris buffer, Mg^{2+}, NaCl, 2-mercaptoethanol, and bovine serum albumin. All endonucleases require Mg^{2+} as a cofactor, and most are active at pH values ranging from 7.2 to 7.6. The major difference among the endonucleases is their dependence on ionic strength and their temperature optima. Most digestions are done at 37°C, but a few restriction endonucleases perform better at lower temperatures. Endonuclease activity is usually measured with bacteriophage λ DNA as a substrate by the manufacturer, but activity of the endonuclease varies greatly with different DNA substrates and also can be modified by the neighboring sequences. Enzyme activity rates can vary by a factor of 10- to 50-fold in your laboratory experiments.

The number and variety of endonucleases available for genetic manipulations continues to increase. Endonucleases that recognize longer recognition sequences are particularly useful if large DNA fragments are to be separated by pulsed-field gel electrophoresis. New microbial sources of enzymes are being sought, especially those that tolerate high temperatures. Catalogues obtained from many suppliers contain useful information on restriction endonuclease activity and their appropriate reaction conditions.

5.7. Joining DNA Molecules

Different DNA fragments cleaved by restriction endonucleases can be joined together by ligases, which makes it possible to insert exogenous DNA into plasmid vectors. Two **DNA ligases** are commonly used. One is derived from *E. coli* and the other from the bacteriophage T4. Their requirements for cofactors differ. T4 ligase requires ATP while *E. coli* ligase requires NAD^+. Both catalyze the joining of a 5′-phosphate and a 3′-OH group to form a phosphodiester bond. T4 DNA ligase will catalyze the joining of blunt-ended DNA molecules and cohesive-ended molecules, although more enzyme is required for blunt-ended ligations.

If the restriction endonuclease used generated DNA fragment ends with uneven ends or overhangs (Table 5.4), then the sequences of the DNA within the single-stranded regions of the two molecules have to be complementary for ligation to occur. Ligation of 4-base extensions is easier than ligation of 2-base extensions. Extensions that consist of G+C bases ligate more readily than those with A+T base pairs (can you explain why?).

Blunt ends are more difficult to ligate, requiring 20 to 100 times more T4 DNA ligase and higher DNA concentrations. The surrounding DNA sequences do not affect ligation efficiency, but ligation is negatively influenced by the presence of contaminating endonucleases or by phosphatase.

Table 5.5. Blunt-End Ligation When the Vector to Insert Molar Ratio Is 3

In a siliconized Eppendorf tube in ice mix:	
Dephosphorylated vector DNA (∼4 kb)	160 ng
DNA fragment (∼1 kb)	13 ng
10 × ligase buffer I (250 mM Tris-HCl at pH 7.5, 50 mM $MgCl_2$, 25% w/v polyethylene glycol 8000, 5 mM DTT, 4 mM ATP)	4 μl
T4 DNA ligase	1 Weiss unit
Water to a final volume of	20 μl
Incubate at 23°C for 4 hr and stop the reaction by adding 1 μl of 0.5 M EDTA. Dilute fivefold before adding the mixture to competent *E. coli* cells for transformation.	

The optimum temperature for ligating DNA is 37°C, but the hydrogen-bonded joint between sticky ends is unstable at this temperature. The ligation reaction is carried out at a temperature which is a compromise between the optimum for the rate of the enzyme action and that for the association of the termini, and usually is done between 4 and 15°C. Ligation reactions often are allowed to take place overnight at these low temperatures.

The ligation reaction can be promoted by adjusting the ratio of insert DNA and vector DNA. Ligation can also be improved when a linear DNA fragment is produced by a restriction endonuclease from a circular vector; the linear fragment will often recircularize and hydrogen bond to itself, or to other linear vector sequences. To prevent this, the linearized plasmid vector DNA can be treated with alkaline phosphatase to remove 5′-terminal phosphate groups. Alkaline phosphatase prevents recircularization of the plasmid or formation of plasmid dimers, although the phosphatase must be eliminated if the vector and linear DNA are to be ligated. Circularization of the vector will then occur if the foreign DNA (untreated with phosphatase) joins the ends of the vector.

Only T4 DNA ligase is able to join blunt-ended DNA molecules. A typical blunt end ligation reaction is described in Table 5.5.

5.8. Growth, Maintenance, and Storage of *E. coli*

Escherichia coli has been studied extensively and its genome has been completely sequenced (Snyder and Champness 1997, Neidhardt 1999). DNA manipulations require manipulating bacteria, primarily derivatives of *E. coli* K12 strains. Different *E. coli* strains are used for different purposes (Miller 1992).

Standard microbial techniques are employed: pure cultures of *E. coli* are obtained by propagating cultures from single, isolated colonies on agar plates. Dilution streaking with an inoculating loop readily produces isolated colonies, and an isolated colony can be restreaked to obtain a pure master plate which can be stored at 4°C for a month. It is important that *E. coli* cultures be kept pure, that the phenotypes be verified prior to use, and that the cultures be stored properly. For long-term storage, cultures can be stored in stab vials, as frozen glycerol cultures, or as lyophilized cultures.

Overnight cultures of most strains of *E. coli* produce $\sim 4 \times 10^9$ bacteria per ml depending upon the medium, the degree of aeration, the strain, and the temperature. To determine the cell concentration, dilutions of the culture should be plated. Detailed methods for manipulating *E. coli* are readily available in a number of laboratory manuals (Miller 1992, Sambrook and Russell 2001).

5.9. Plasmids for Cloning in *E. coli*

Plasmids are widely used as cloning vectors. Many plasmids have undergone extensive genetic engineering to enhance their value as vectors (Figure 5.2). The complete sequence of the vector usually is known, including the location of unique restriction sites (sites where a specific endonuclease can cut the plasmid).

Wild-type **plasmids** are small DNA molecules that are stably inherited as extrachromosomal units in many kinds of bacteria. Plasmids are widely found in bacteria, but usually are not essential to their host. Many plasmids carry genes for antibiotic resistance, antibiotic production, heavy metal resistance, an ability to degrade aromatic compounds, sugar fermentation, enterotoxin production, or hydrogen sulfide production. Most are covalently closed DNA circles, but some are linear.

Plasmids can be classified into two types depending upon whether they carry a set of genes that promote **bacterial conjugation**. Plasmids also can be categorized as to whether

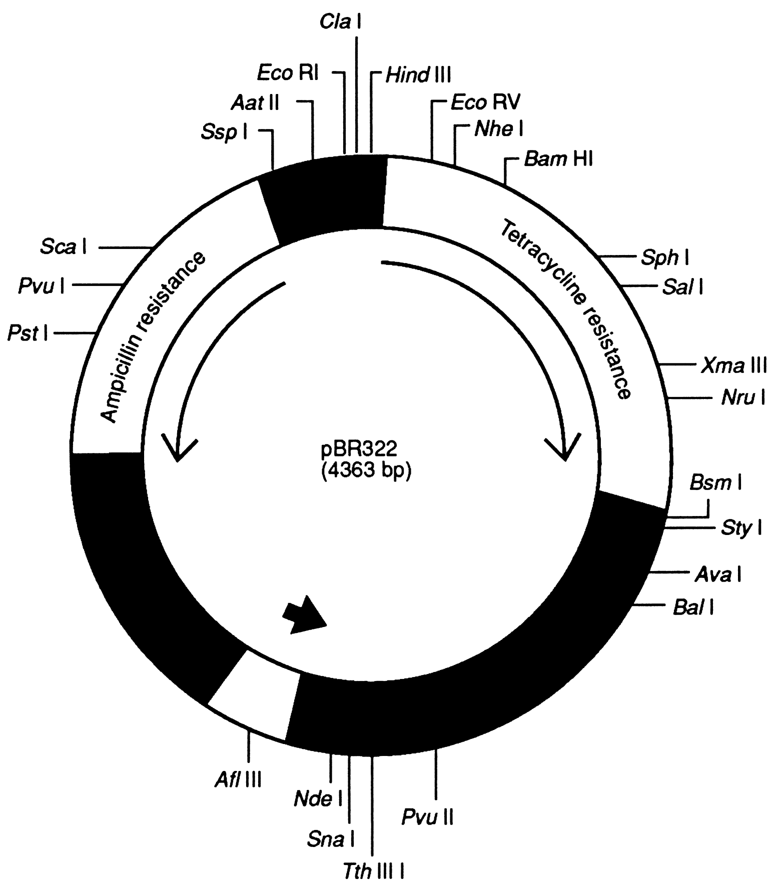


Figure 5.2. The structure of the cloning plasmid pBR322 showing the unique sites where restriction endonucleases can cleave the DNA. The thin arrows inside the circle indicate the direction of transcription of the ampicillin and tetracycline resistance genes, which serve as selectable markers. The thick arrow shows the direction of DNA replication.

they are maintained in multiple copies in host cells or in limited numbers per cell. Generally, plasmids that promote bacterial conjugation are relatively large and are present in one to three copies per bacterial cell. Plasmids that do not promote bacterial conjugation are smaller, and multiple copies are found in a cell. Plasmids are “promiscuous” if they are able to promote their own transfer to a wide range of bacteria and can be maintained stably in their new hosts. Promiscuous plasmids can transfer cloned DNA molecules into different bacteria. Wild-type plasmids could be used for cloning in *E. coli*, but they suffer from a number of disadvantages, and genetically engineered plasmids have been developed that have a number of desirable attributes.

The first genetic improvement of plasmids involved removing excess DNA so that the plasmid is easier to manipulate *in vitro*, resistant to damage by shearing, and readily isolated from bacterial cells. Smaller size is an advantage because bacterial cells usually can sustain several smaller plasmids, which will increase the yield of the recombinant DNA molecules.

A second improvement was the addition of one or more **selectable marker** genes to the plasmid (Figure 5.2). A selectable marker allows the experimenter to identify those bacterial cells that have taken up the plasmid during the transformation process. Many selectable markers are antibiotic resistance genes (for example, ampicillin and tetracycline) that allow the transformed bacteria to be grown on selective media.

A third improvement involved adding DNA containing a **multiple cloning site** or **polylinker** that can be cut by several restriction endonucleases (Figure 5.3). The presence of these unique restriction or cloning sites is helpful because cloning requires that both the vector and the exogenous DNA be cut with the same endonuclease (or ones that produce the same kinds of ends) so that the ends can be ligated together. If the plasmid had more than one site that was cut by a specific endonuclease, the plasmid vector would be cut into several fragments, resulting in defective vectors. A polylinker is a short segment of DNA with sites where several different restriction endonucleases can cut. This gives the genetic engineer options as to which restriction enzyme to use. If the polylinker is placed within a selectable marker gene, such as *lacZ*, the gene function is disrupted when exogenous DNA is cloned into it at any of the restriction sites, and the recombinant colonies can be identified by their color.

Plasmids have been engineered by sophisticated techniques to perform a variety of defined tasks. **Expression vectors** facilitate expression of proteins; for example, baculovirus vectors are used to produce large amounts of foreign proteins in insect cells (Chapter 6). Some vectors help identify regulatory signals that turn genes on or off, some are used for direct selection of recombinants, some have increased stability so that they are not eliminated from their host cells, and others are genetically altered so that high copy numbers per host cell can be maintained. **Sequencing vectors** are particularly useful for DNA sequencing because they produce a single-stranded copy of DNA rather than a double-stranded molecule (Chapter 7). Some vectors are modified so that proteins are secreted through the host cell wall which facilitates purification of proteins, and others are modified to produce fusion proteins to facilitate protein purification. Many versatile vectors are available from commercial sources.

A key to effective genetic engineering is the ability to identify those cells that have been genetically transformed. To aid in this, most plasmid vectors contain at least two selectable markers, which often are antibiotic resistance genes. If two markers are present, then one is often the site into which the exogenous DNA is cloned. Insertion of exogenous DNA should inactivate that resistance gene, so that the *E. coli* cells containing this insertion can be identified because they are newly susceptible to the antibiotic. Another commonly

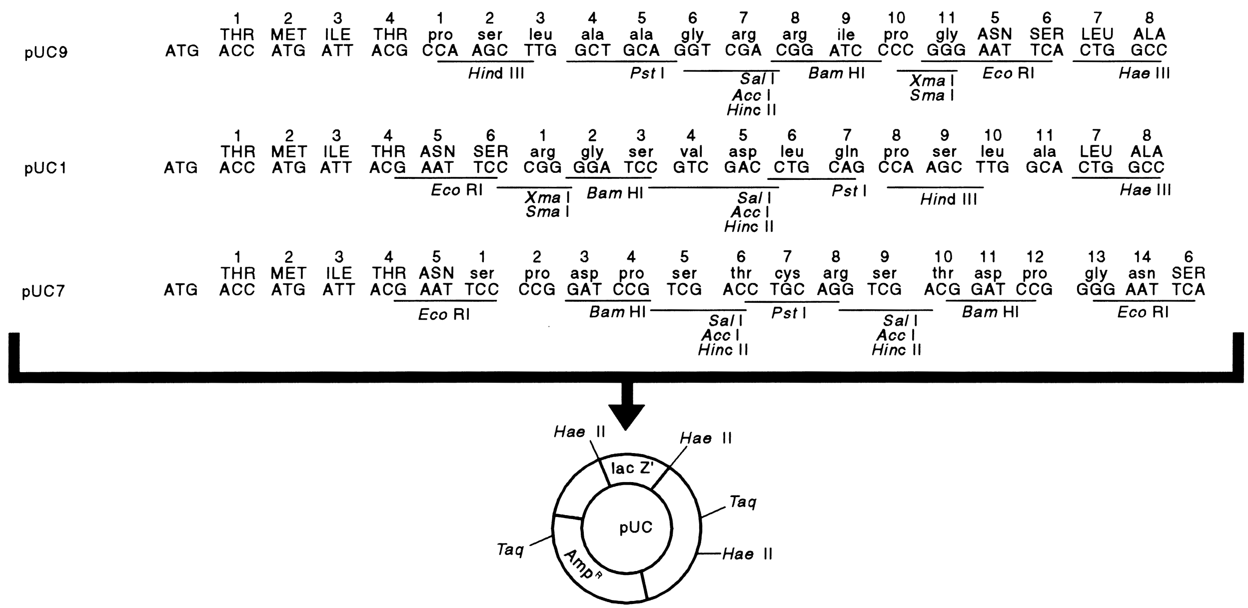


Figure 5.3. Genetic maps of the multiple cloning sites of some pUC plasmids. The multiple cloning site is inserted in the *lacZ* gene but does not disrupt *lacZ* gene function. If exogenous DNA is cloned into this site, then *lacZ* gene function is disrupted, which indicates that insertion was successful. This allows the genetic engineer to detect colonies of *E. coli* containing recombinant plasmids because these colonies are white rather than blue. The pUC plasmids also contain an ampicillin resistance gene.

used selectable marker is *lacZ*, which allows selection of blue/white bacterial colonies. This plasmid is described in more detail in Chapter 6.

Plasmid vectors, such as pBR322 and its derivatives, are widely used because they can be produced in multiple copies within a cell, they are easily purified, and they can produce large amounts of the cloned gene (Figure 5.2). pBR322 carries both ampicillin and tetracycline resistance genes and an **origin of replication**, which is a sequence at which replication of the DNA molecule is initiated. pBR322 has been completely sequenced and its restriction sites totally characterized. This means that the exact length of each fragment from a restriction digest can be predicted and these fragments can serve as DNA markers for sizing other DNA fragments. pBR322 fragments produced after digestion with restriction enzymes range in size from several base pairs to the entire 4.3-kb-long plasmid.

Some unique restriction sites occur within both the ampicillin and tetracycline resistance genes of pBR322 (Figure 5.2) and these are very useful in cloning. If exogenous DNA is inserted into a site in the ampicillin resistance gene where a restriction enzyme cuts uniquely, the ampicillin resistance gene will be inactivated. The recombinant plasmids contained within their *E. coli* host can then be identified by first spreading the transformed *E. coli* onto culture plates with media containing tetracycline and then replica plating them onto plates with ampicillin.

Replica plating is a procedure in which a particular pattern of *E. coli* colonies on an agar surface is reproduced on another agar surface. The pattern is obtained by pressing a piece of sterile velvet upon the original agar surface, which transfers cells from each colony to the cloth, and then pressing this pattern onto another agar surface. It is important to carefully mark the orientation of the patterns on the original and replica plates. Those recombinant *E. coli* colonies that are unable to survive on the ampicillin can be recovered by finding the colony growing on the original tetracycline plate based on their location. Many derivatives of pBR322 have been constructed to fulfill particular cloning goals (Balbas et al. 1986).

5.10. Transforming *E. coli* with Plasmids

A plasmid carrying exogenous DNA must be inserted into *E. coli* in order to amplify, or **clone**, the DNA. The process of inserting a plasmid into *E. coli* is called bacterial **transformation**. For many years, efforts to transform *E. coli* were unsuccessful, and it was only in 1970 that methods were developed. The ability to transform *E. coli* required understanding its genetics and having the ability to manipulate the physiological status of the *E. coli* cells to optimize the transformation reaction.

A simple transformation procedure involves suspending *E. coli* cells that are in the log phase of their growth cycle in an ice-cold solution containing membrane-disrupting agents such as PEG, DMSO, or divalent cations such as calcium chloride (Table 5.6). Plasmid DNA is then added to a small aliquot of these **competent cells** (competent for transformation), and the incubation on ice is continued for another 30 min. A heat shock is then administered by putting the cells into 42°C for 2 min. The cells are then transferred to nutrient broth and incubated for 30 to 60 min to allow the plasmid to express its phenotypic properties (plasmids often carry antibiotic resistance genes as selectable markers). The cells then are plated onto agar plates containing a selective medium. Only those bacteria that have taken up the plasmid with the selectable marker should survive and reproduce on the selective medium.

Table 5.6. Producing Competent *E. coli* and Transforming Them with Plasmid DNA Using $CaCl_2$

Producing competent cells

1. Grow a fresh overnight culture of *E. coli* in LB broth at 37°C.
2. Dilute the cells 40-fold into 1 liter of fresh medium. Incubate at 37°C with good aeration until their density produces an absorbency rating at 550 nm of 0.4–0.5.
3. Immediately chill the culture by swirling in a ice-water bath.
4. When the cells are chilled, centrifuge the culture at 4°C at 5000 rpm for 10 min.
5. Decant the supernatant and place the pellet in ice.
6. Resuspend the pellet in 500 ml ice-cold 100 mM $CaCl_2$. It is easier to resuspend the pellets if they are vortexed before the $CaCl_2$ is added. The cells can be suspended by sucking them up and down in a 25-ml pipette.
7. Once the cells are resuspended, incubate in ice for 30 min with occasional swirling.
8. Pellet the cells once again at 5000 rpm for 10 min at 4°C.
9. Resuspend in 40 ml of ice-cold 100 mM $CaCl_2$ and 15% glycerol.
10. Distribute aliquots of 0.2 ml of cells into sterile Eppendorf tubes in ice.
11. Keep in ice at 0–4°C for 12–24 hr. This is essential for maximal competency, although the cells are competent at this stage.
12. Freeze the tubes in ethanol–dry ice or liquid nitrogen and place immediately at −70°C. The cells remain competent for months if stored at −70°C.

Transformation of competent cells

1. Thaw a tube of frozen competent cells at 4°C.
2. Add DNA in buffer.
3. Incubate in ice for 30 min.
4. Heat shock for 2 to 5 min in a 42°C water bath.
5. Add 0.4 ml of LB broth at room temperature to each tube and incubate for 1 hr at 37°C.
6. Spread on agar plate with appropriate antibiotics. Incubate plates overnight at 37°C.

Modified from Berger and Kimmel (1987).

How transformation occurs is not entirely understood. Various agents affect the bacterial cell wall and, in the case of $CaCl_2$, also may be responsible for binding DNA (the plasmid) to the cell wall. The actual uptake of DNA is stimulated by the brief heat shock. Large DNA molecules are taken up less efficiently than smaller ones; the efficiency of transformation varies with the strain of *E. coli* used and is typically expressed as the number of transformant cells per μg of plasmid DNA. Various protocols produce efficiencies of 10^7 or 10^8 transformants/μg of plasmid DNA.

Electroporation can be used to insert DNA into bacterial cells as well. Electroporation involves disrupting the cell membrane briefly by an electrical current so that DNA can be incorporated. Commercial units and protocols can be purchased for electroporation of *E. coli*.

5.11. Purifying Plasmid DNA from *E. coli*

Removing the plasmids from *E. coli* is necessary if experiments are to be conducted on the now-cloned DNA. The trick is to lyse (break open) the *E. coli* cells just sufficiently so the plasmids can escape without too much contamination by the bacterial chromosome. If the bacterial cell is lysed gently, most of the bacterial chromosomal material released will be of higher molecular weight than the plasmids and can be removed, along with the cell

debris, by complexing with detergents and high-speed centrifugation. The plasmid DNA, which is left in the clear liquid remaining, then can be extracted by one of two traditional methods.

1. Cesium chloride centrifugation with **ethidium bromide** (EtBr) yields bands in the centrifuge tube that contain chromosomal and plasmid DNA at different levels due to the different densities of linear and supercoiled DNA in the presence of EtBr.

Ethidium bromide stains DNA by intercalating between the double-stranded DNA base pairs and in so doing causes the DNA to unwind (Figure 5.4). A plasmid DNA molecule that has not been nicked is a circular double-stranded supercoil which has no free ends and can only unwind to a limited extent, thus limiting the amount of EtBr that it binds. A linear DNA molecule, such as fragmented bacterial chromosomal DNA, can bind more EtBr and become stiffer, extending the molecule and reducing its buoyant density.

2. A second method for extracting and purifying plasmid DNA exploits the observation that within a pH range of 12.0 to 12.5, linear DNA will completely denature, but closed

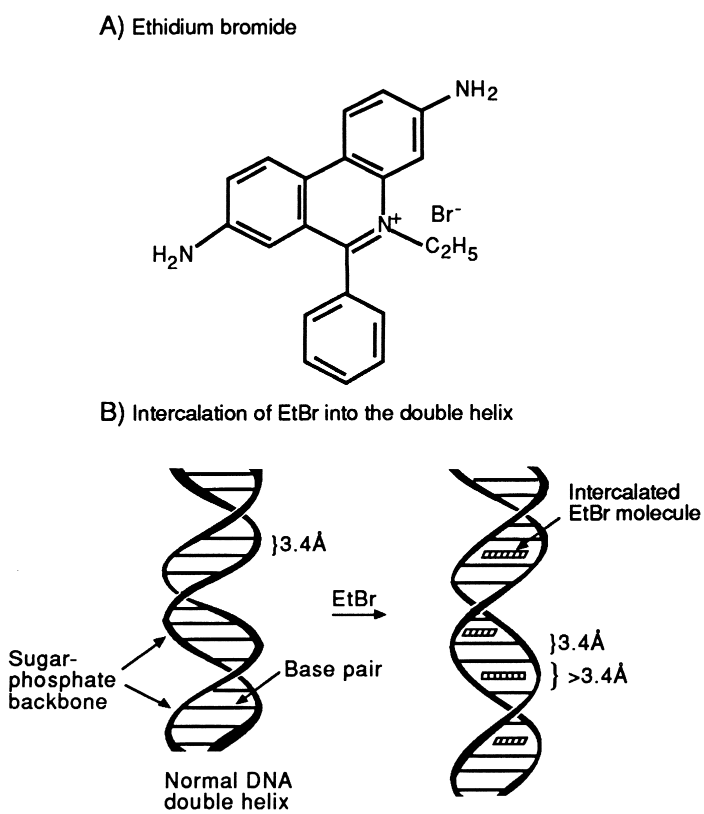


Figure 5.4. Ethidium bromide (EtBr) can intercalate into DNA and cause the DNA to unwind. DNA containing EtBr will fluoresce if exposed to UV radiation. EtBr is used to visualize DNA fragments on gels after electrophoresis. EtBr can also serve as a mutagen.

Table 5.7. Small-Scale Plasmid Preparations (Minipreps)

1. Prepare 5-ml transformed *E. coli* cultures in LB broth containing the appropriate antibiotic. The cultures can be grown in disposable 14-ml plastic centrifuge tubes by picking colonies with a sterile toothpick and dropping the toothpick into the tube. Cap the tube and incubate at 37°C with shaking (250 rpm) for 16 hr.
2. Pellet the cells by centrifugation at 5000 rpm for 5 min. Discard supernatant and toothpick.
3. Add 100 μl of 50 mM glucose, 25 mM Tris-HCl (pH 8), 10 mM EDTA, 2 mg/ml lysozyme (freshly prepared). Resuspend pellet and incubate for 10 min.
4. Add 200 μl 0.2 N NaOH, 1% SDS. Mix gently. Incubate on ice for 10 min. The SDS–NaOH solution must be made just prior to use and kept at room temperature. Mix 3.5 ml water, 1 ml 1 N NaOH, and 0.5 ml 10% SDS.
5. Add 150 μl 3 M potassium acetate (pH 4.8). Mix gently. Incubate for 10 min in a freezer. A white precipitate will form.
6. Centrifuge for 15 min at 15,000 rpm at 4°C.
7. Pour supernatant into Eppendorf tubes and fill with cold ethanol. Incubate in ice for 10 min.
8. Pellet the DNA for 1 min in a microcentrifuge and aspirate off supernatant. Add 0.5 ml cold 70% ethanol and aspirate off.
9. Dry under vacuum. Resuspend in 50 μl distilled water containing 10 μg/ml pancreatic ribonuclease (RNase) to remove RNA.
10. The DNA is suitable for restriction analysis or fragment preparation. Use 5–10 μl per reaction. If the DNA does not cut well, it can be re-extracted with phenol and precipitated with ethanol.

circular (plasmid) DNA will not. Plasmid-containing bacteria are treated with lysozyme to weaken the cell wall and then lysed with sodium hydroxide and SDS (sodium dodecyl sulfate). The chromosomal DNA is denatured, but upon neutralization with acidic potassium acetate, the chromosomal DNA renatures and aggregates to form an insoluble network. The high concentration of potassium acetate also causes the protein–SDS complexes and high molecular RNA to precipitate. If the pH of the alkaline denaturation step has been controlled carefully, the plasmid DNA molecules will remain circularized and in solution while the contaminating molecules precipitate. The precipitate can be removed by centrifugation and the plasmid purified and concentrated by ethanol precipitation (Table 5.7).

Many kits are now available that can be used to extract plasmids from transformed *E. coli*, and automated work stations can be purchased for laboratories with a high-volume work load.

5.12. Electrophoresis in Agarose and Acrylamide Gels

DNA and RNA molecules can be separated by size and visualized by agarose or acrylamide gel electrophoresis. **Gel electrophoresis** provides a powerful method for resolving mixtures of single- or double-stranded (ds) nucleic acid molecules. The nucleic acids can be visualized *in situ* in the gel by staining with EtBr and examining the gel under UV light (Figure 5.5).

At a pH near neutrality, linear DNA is negatively charged and migrates from cathode to anode in a gel, with its mobility dependent on fragment size, voltage applied, composition of the electrophoresis buffer, base composition, gel concentration, and temperature.

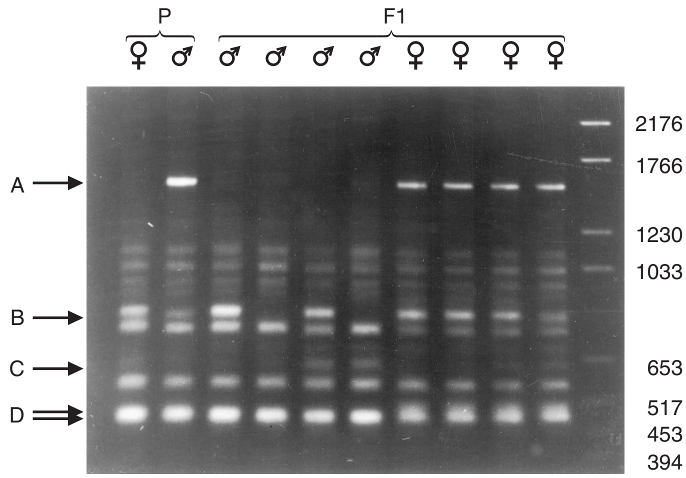


Figure 5.5. Photograph of DNA stained with ethidium bromide and illuminated with UV light. The bands are a pink-purple color.

Agarose gel is used for longer DNA molecules and **polyacrylamide** for shorter. Nondenaturing polyacrylamide gels can be used to separate double-stranded DNA fragments between 6 bp (20% acrylamide) and 1000 bp (3% acrylamide) in length. Nondenaturing agarose gels can separate DNA fragments between 70 bp (3% agarose) and 10,000 bp (0.1% agarose). Single-stranded DNA can be separated by agarose or polyacrylamide gel electrophoresis by including a denaturing reagent in the gel.

DNA from 60 kb to 0.1 kb can be detected with agarose gels containing different percentages of agarose. Thus, it is possible to separate DNA that is 0.1 to 3 kb long in 2% gels, DNA that is 0.8 to 10 kb long in 0.7% gels, and DNA that is 5 to 60 kb long in 0.3% gels. Agarose gels are usually electrophoresed at room temperature, except for low-percentage agarose gels (<0.5%) which are easier to handle at cooler temperatures, and low-melting-temperature agarose gels, which may melt if run too fast at room temperature.

Agarose powder comes in grades which vary in purity and melting temperature. Type II agarose is generally used, although it contains contaminants that coelute with DNA and inhibit most commonly used enzymes. This means that DNA must be purified following elution from this gel if it is to be ligated or cut with restriction enzymes. An alternative involves using high-quality, low-melting temperature agarose which melts at 65°C and sets at 30°C. Low-melt agarose allows DNA to remain double-stranded and also allows many enzymes to be used in the liquid agar.

Polyacrylamide gels result from polymerization of acrylamide monomers into linear chains and the linking of these chains with *N*,*N*′-methylenebisacrylamide (often called bis). The concentration of acrylamide and the ratio of acrylamide to bis determine the pore size of the resultant three-dimensional network and thus its sieving effect on nucleic acids. Polyacrylamide gels can be used to purify synthetic oligonucleotides, isolate or analyze DNA less than 1 kb in size, and resolve small RNA molecules by two-dimensional or pulsed-field gel electrophoresis. Polyacrylamide gels also can be used to sequence DNA. Polyacrylamide gel ingredients are highly toxic and should not be inhaled or touched unless wearing gloves.

5.13. Detecting, Viewing, and Photographing Nucleic Acids in Gels

Ethidium bromide is a useful dye to detect both single- and double-stranded nucleic acids in both agarose and polyacrylamide gels (Figure 5.5). Agarose gels are somewhat less sensitive in detecting small amounts of DNA than are polyacrylamide gels. The sensitivity for single-stranded DNA is 5- to 10-fold less.

Ethidium bromide (EtBr) can be incorporated into the gel and running buffer during electrophoresis. Alternatively, gels can be stained after electrophoresis by placing them in buffer containing EtBr for 30 min. Ethidium bromide is mutagenic, and thus the experimenter must be extremely cautious when handling it. Gloves must be worn, and care must be taken to avoid contaminating laboratory surfaces.

As little as 0.05 μg of DNA can be visualized in one band when the gel is exposed to ultraviolet (UV) light (Figure 5.5). The EtBr–nucleic acid complex absorbs UV irradiation at about 260 nanometers (nm) or 300 nm. The fluorescence of EtBr stacking in duplex DNA is 10 times greater than that of free EtBr and the emission is at 590 nm (red orange). UV light sources (transilluminators) are used to view DNA stained with EtBr at 254 or 366 nm. Although the short-wave model can detect smaller amounts of sample DNA, it damages DNA by nicking or dimerization, making it unsuitable for most cloning work. Safety glasses and a face mask must be worn by the experimenter around these UV light sources.

Agarose and polyacrylamide gels can be dried by heating under vacuum in dryers, and the dried gels can be stored for reference. Drying must be performed prior to autoradiography, and agarose gels should not be heated above their melting temperature. Usually, however, photographs are taken of the gels to document the results of the experiment.

5.14. Identifying DNA by Southern Blot Analysis

It is often necessary to identify specific DNA sequences. One method to do so was invented by Southern (1975), and the "Southern blot" has been a fundamental and versatile tool for genetic engineers ever since. This original methods paper by Southern has been the most cited paper ever published in the *Journal of Molecular Biology* (Southern 2000).

Variations on **Southern blots** have been developed to identify RNA or proteins in gels. These modifications are called **Northern** (RNA) and **Western** (protein) **blotting**, respectively.

In Southern blotting, DNA fragments that have been separated by electrophoresis in an agarose gel are denatured into a single-stranded form by alkali treatment (Figure 5.6). The gel is then laid on top of buffer-saturated filter paper. The top of the gel is covered with a nitrocellulose filter membrane. This membrane is then overlain with dry filter paper. Additional layers of dry filter paper or absorbent papers are stacked on top. Buffer passes out of the gel, drawn by capillary action into the dry filter papers. As the buffer moves from the gel, it elutes some of the denatured DNA from the gel. When the single-stranded DNA comes in contact with the nitrocellulose lying on top of the gel, it binds to the nitrocellulose membrane. The blotting process is carried out over several hours and results in the transfer of part of the DNA from the gel onto the nitrocellulose membrane. It results in an (at this point) invisible pattern of bands on the membrane surface that resembles the original bands in the gel, with a minimal loss of resolution. The stack of filter papers is then removed and the nitrocellulose membrane is baked at 80°C in a vacuum to bind the DNA

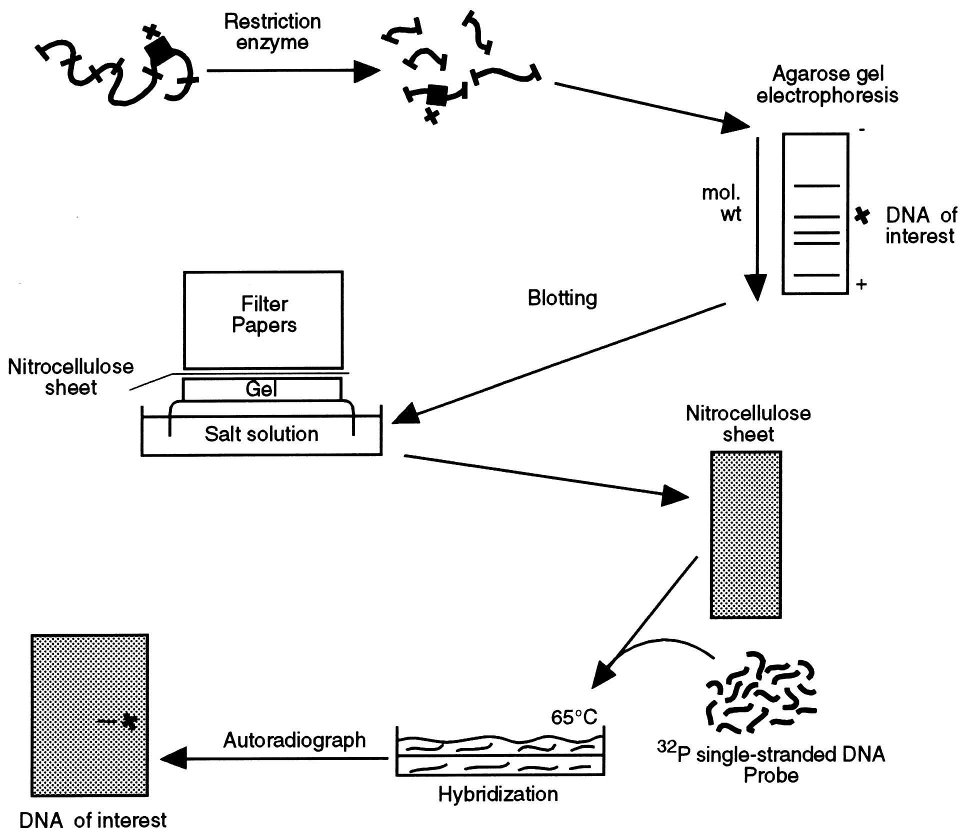


Figure 5.6. Outline of a Southern blot procedure. DNA is cut with restriction enzymes, electrophoresed, and blotted onto nitrocellulose by capillary action. The nitrocellulose sheet containing the DNA is baked to bind the DNA to the nitrocellulose. Specific DNA is identified by the binding of a labeled probe (here ^{32}P-labeled single-stranded DNA) in a hybridization procedure. Excess probe is washed off, and the nitrocellulose sheet is then exposed to X-ray film to visualize the location of the DNA that contains sequences homologous to the radiolabeled probe. The conditions (stringency) under which hybridization occurs can be varied to increase or decrease the specificity of the reaction between DNA and probe. Southern blots allow the genetic engineer to locate specific DNA sequences.

permanently onto the surface of the nitrocellulose. Alternative methods for transferring the DNA to a membrane involve electroblotting or vacuum blotting, which requires specialized equipment.

To determine whether the DNA of interest is present on the blot requires probing the DNA on the nitrocellulose membrane (Figure 5.6). The **probes** can be labeled by radiolabeled ^{32}P or by nonradioactive methods. Probes can consist of radiolabeled RNA, single-stranded DNA, or a synthetic oligonucleotide which is complementary in sequence to the DNA of interest. The labeled probe must bind specifically to the DNA of interest but not bind to the nontarget DNA or the nitrocellulose.

To prevent nonspecific binding, especially by single-stranded DNA probes which have a high affinity for nitrocellulose, the nitrocellulose with the bound DNA of interest is pretreated by placing it in a solution containing 0.2% each of Ficoll (an artificial polymer

of sucrose), polyvinylpyrrolidone, and bovine serum albumin (also known as Denhardt's solution). The mixture often includes an irrelevant nucleic acid such as salmon sperm DNA, which may act by occupying all the available nonspecific binding sites on the membrane.

The temperature at which Southern blotting is conducted is adjusted to maximize the rate of hybridization of the probe with the immobilized DNA on the nitrocellulose and to minimize the amount of nonspecific binding. This aspect of planning the Southern blot is called **stringency**, and a highly stringent Southern blot will be more specific. After the hybridization step, in which the labeled probe binds to the immobilized DNA on the membrane, the membrane is washed to remove any unbound probe. The temperature at which the washing takes place also determines the stringency of the Southern blot.

The regions on the membrane where hybridization of the labeled probe and target DNA took place are detected by placing the membrane in contact with X-ray film if the probe was radiolabeled. The length of time the X-ray film is exposed to the radioactivity is determined by the amount of DNA in the blot and the degree of homology between the DNA and the probe. If there is only a small amount of DNA present, as would be expected for a single-copy gene in a blot of genomic DNA, the film may be exposed for 2 to 3 weeks to the blot.

Modifications of the Southern blot method now employ nylon membranes as substrates because they are more robust and can be reused. Thus, one probe can be removed by high-temperature washing, and the DNA can be reprobed with a different probe. Another advantage to nylon membranes is that the DNA can be permanently fixed to the membrane by a brief exposure to UV light, which cross-links the DNA and fixes it to the membrane.

5.15. Labeling DNA or RNA Probes

In molecular biology, many DNA manipulation techniques depend on hybridizing a nucleic acid probe to a target DNA or RNA sequence. Probes are required in Southern and Northern blots, dot blots, colony/plaque blots, and *in situ* hybridization. Dot blots can be used to identify unfractionated DNA or RNA molecules that have been immobilized on a nitrocellulose membrane. Plaque/colony blots detect DNA released from lysed bacteria or phage after immobilizing the DNA on a nitrocellulose membrane. *In situ* hybridization is employed to detect DNA or RNA molecules in cytological preparations.

Nucleic acid probes can be labeled by several methods, as will be described in Chapter 6. One method to uniformly label double-stranded DNA probes is described here. **Nick translation** describes the incorporation of a nick (or break in one strand) of a double-stranded DNA molecule (Figure 5.7). Nicks are introduced at widely separated, random sites along the DNA molecule by treating the DNA with small amounts of DNase. A nick exposes a free 3′-OH group, and DNA polymerase I of *E. coli* will then remove nucleotides from the 5′ side of the nick. The simultaneous removal of nucleotides from the 5′ side and the addition of labeled nucleotides to the 3′ side by DNA polymerase I result. If dNTPs are radiolabeled with ^{32}P and the nicks are random, the duplex DNA molecule will become labeled uniformly along its length as it incorporates radiolabeled dNTPs. The reaction may be carried out to label all four dNTPs or only one dNTP. Nick translation is particularly useful for producing large amounts of probe for multiple hybridization reactions and where a high probe concentration is required.

Nick translation kits are available from a number of commercial sources and provide instructions, a stock mixture of DNA polymerase I and DNase I, and a series of buffers

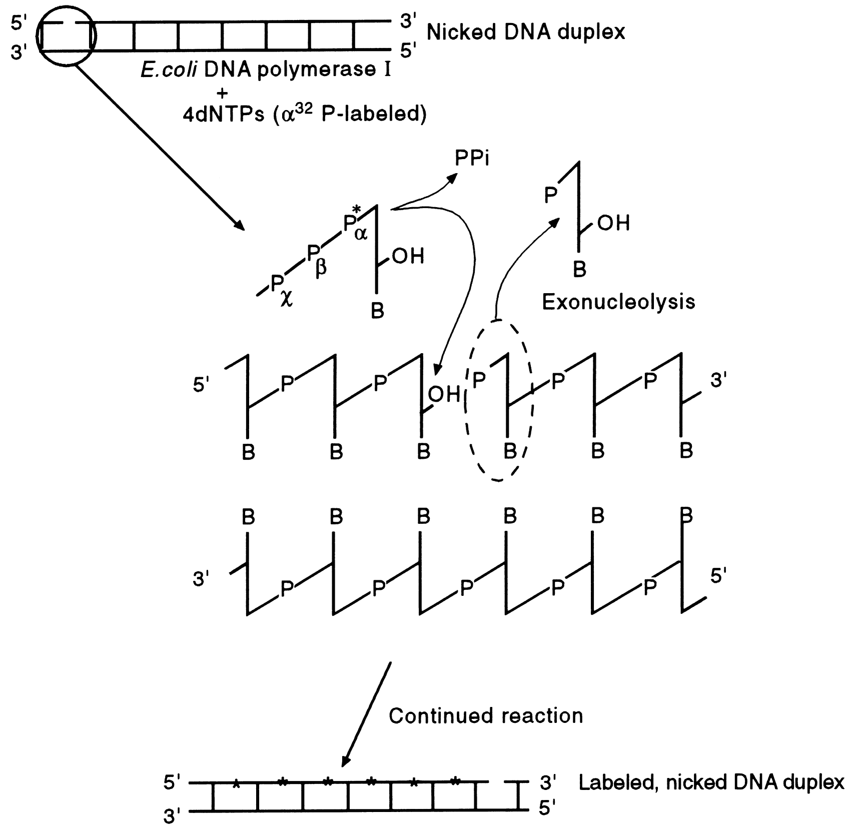


Figure 5.7. ^{32}P-labeling of double-stranded DNA can be carried out by nick translation. The asterisks indicate the location of radiolabeled phosphate groups that are inserted into the strand.

lacking one or more unlabeled dNTPs. The radioactive dNTPs must be obtained fresh (within a few days) before the labeling reaction is set up because ^{32}P-labeled dNTPs decay rapidly.

The use of radioactive probes requires that the experimenter obtain training in their safe use. Safety protocols also are required to prevent contamination of the laboratory and to ensure safe disposal of the radioactive dNTPs. Most organizations require specialized training and licensing of scientists using radioactivity and reports to confirm its safe use and disposal (Davies 1994).

Kits are now available to label DNA without using radioactivity, which reduces potential risks to the experimenter and eliminates the need for special disposal methods.

5.16. Removing DNA from Agarose Gels after Electrophoresis

Several methods have been developed to recover DNA from agarose gels, including: 1) electrophoresis onto a DEAE-cellulose membrane, 2) electroelution into dialysis bags, and 3) use of low-melting temperature agarose gels. With method 1, fragments of DNA are

separated by electrophoresis, a slit is cut in the gel immediately ahead of the DNA fragment of interest, and a sliver of DEAE-cellulose membrane is inserted. Electrophoresis is continued until the DNA in the band has been transferred to the membrane. The membrane is removed from the slit and washed, and the DNA is eluted from the membrane.

Method 2, electroelution, allows recovery of a wide size range of DNA but is inconvenient. DNA is separated by electrophoresis in agarose gel containing EtBr. The band of interest is located with an ultraviolet lamp, and the band is cut from the gel with a razor blade. The gel fragment containing the DNA of interest is then placed in a piece of dialysis tubing, sealed, and placed into an electrophoresis tank. Electric current is passed through the bag to elute the DNA out of the gel and onto the inner wall of the bag. The polarity of the current is reversed to release the DNA from the wall of the bag, the bag is opened, and all the buffer containing the DNA is transferred to a clean tube.

Method 3 uses low-melting-temperature agarose gels. DNA of interest is electrophoresed, and the band of interest is detected by staining with ethidium bromide, cut out, and placed into a clean tube. A buffer is added, and the mixture is heated to 65°C to melt the agarose, then centrifuged, which will leave the DNA in the aqueous phase and the agarose in the interface. The DNA in the aqueous phase can be purified with phenol. The DNA in the aqueous phase is then precipitated by ammonium acetate and cold ethanol. At this point the DNA is sufficiently pure to be digested by restriction enzymes or modified by ligases.

None of these methods is fully satisfactory in producing large amounts of pure DNA. Problems include the presence of inhibitors of enzymatic inhibitors in the agarose, which can affect subsequent DNA manipulations. Large fragments (>5 kb) of DNA are often inefficiently recovered from agarose gels because these longer fragments bind very tenaciously to purification matrices such as the DEAE-cellulose membrane. Small (<500 nanograms or ng) amounts of DNA are recovered inefficiently from gels. The methods are labor-intensive so that recovery of DNA from gels is not readily performed on large numbers of samples. Various kits can be purchased from commercial sources to purify DNA from gels.

5.17. Restriction Site Mapping

So far in this experiment, DNA has been cloned into a plasmid and amplified in *E. coli*, and specific sequence(s) have been identified by Southern blot analysis using a radiolabeled probe. Specific bands of DNA have been isolated and the DNA purified from the agarose gel.

Information about the cloned DNA now can be obtained by restriction site mapping, DNA sequencing, or translation of the DNA into proteins using an expression vector. DNA sequencing is described in Chapter 7, and translation and expression of DNA are discussed in Chapter 6.

Restriction site mapping is a relatively simple technique that provides a physical map of the sites in the DNA at which different restriction enzymes cut. One method for constructing restriction maps involves digesting the DNA with a series of single restriction endonucleases in separate reactions. The products of each digestion are electrophoresed on agarose or polyacrylamide gels. DNA marker fragments of known size (size markers) are electrophoresed in lanes adjacent to the DNA being examined to provide estimates of the lengths of the DNA fragments generated. DNA molecular markers of known size are available from commercial sources. The DNA is stained with EtBr, and the bands that were produced are examined under UV light. The banding patterns can be photographed.

After the single digestions are done, the DNA can be digested simultaneously with two restriction enzymes. Again, the sizes of the digestion products are analyzed by gel electrophoresis, using not only size markers but the samples of the first digest for comparison. If the products of the first and second digests are electrophoresed in adjacent lanes on the gel, it is possible to detect small differences in migration rate. Maps are built up from these data by a process of trial and error and basic logic. Based on the sizes of the DNA fragments generated, it is possible to define the order of the restriction sites and the distances between them. The resolution of map distances depends on the accuracy with which the sizes of the DNA fragments can be estimated relative to those of the size markers. However, restriction maps are rarely accurate to less than 100 to 200 bp.

Restriction maps of DNA provide the experimenter with useful information for additional experiments. Furthermore, such restriction site maps can be used as in systematics or population genetics studies (Chapters 12, 13). You will use many of the techniques described in this chapter for other purposes, including preparing a genomic library, as described in Chapter 6.

General References

Ausubel, F. M., R. Brent, R. E. Kingston, D. M. Moore, J. G. Seldman, J. A. Smith, and K. Struhl, Eds. *Current Protocols in Molecular Biology*. Updated since (1987). Wiley Publ., New York. These large red volumes offer research protocols. The publisher has a subscription update service so that new pages can be inserted into the binder to replace outdated material or expand on existing information.

Barker, K. (1998). *At the Bench: A Laboratory Navigator*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. This book is a handbook for beginners in the laboratory. It provides information on making reagents and buffers, their storage and disposal, working without contamination, handling radioactivity, carrying out centrifugation and electrophoresis, using the light microscope, and general laboratory organization and procedures.

Berger, S. L., and A.. R. Kimmel, Eds. (1987). *Guide to Molecular Cloning Techniques*. Methods in Enzymology. Vol. 152. Academic Press, San Diego.

BioSupplyNet Source Book (2001). *The Directory of Research Products and Services for: Molecular Genetics, Cell Biology, Immunology, Biochemistry, High Throughput Analysis*, 7th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Brown, T. A., Ed. (1991). *Essential Molecular Biology. A Practical Approach*. Vols. I and II. IRL Press, Oxford, UK.

Brown, T. A., Ed. (1991). *Molecular Biology LabFax*. Bios Sci. Publ., Oxford, UK.

Glover, D. M. (1984). *Gene Cloning. The Mechanisms of DNA Manipulation*. Chapman and Hall, London, UK.

Hames, B. D. H., and S. J. Higgins, Eds. (1985). *Nucleic Acid Hybridisation—A Practical Approach*. IRL Press, Oxford, UK.

Miller, J. (1992). *A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for* Escherichia coli *and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Old, R. W., and S. B. Primrose (1989). *Principles of Gene Manipulation. An Introduction to Genetic Engineering*, 4th ed. Blackwell Scientific, Oxford, UK.

Sambrook, J., and D. W. Russell (2001). *Molecular Cloning. A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Three volumes provide detailed protocols of methods.

Sambrook, J., E. F. Fritsch, and T. Maniatis (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Three volumes provide detailed protocols of methods.

Sealey, P. G., and E. M. Southern, Eds. (1982). *Gel Electrophoresis—A Practical Approach*. IRL Press, Washington, DC.

Streips, U. N., and R. E. Yasbin, Eds. (2002). *Modern Microbial Genetics*, 2nd ed. Wiley Publ., New York.

Sullivan, W., M. Ashburner, and R. S. Hawley, Eds. (2000). *Drosophila Protocols*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Williams, J. G., and R. K. Patient (1988). *Genetic Engineering*. IRL Press, Oxford, UK.

References Cited

Balbas, P., X. Soberon, E. Merino, M. Zurita, H. Lomeli, F. Valle, N. Flores, and F. Bolivar (1986). Plasmid vector pBR322 and its special-purpose derivatives—a review. *Gene* **50**: 3–40.

Berger, S. L., and A. R. Kimmel, Eds. (1987). *Guide to Molecular Cloning Techniques*. Methods in Enzymology. Vol. 152. Academic Press, San Diego.

Davies, R. W. (1994). Working safely with radioactivity. pp. 471–477. In: *Cell Biology: A Laboratory Handbook*. J. E. Celis, Ed., Academic Press, San Diego.

Frank, S. A. (1994). Polymorphism of bacterial restriction–modification systems: the advantage of diversity. *Evolution* **48**: 1470–1477.

Gilbert, W. (1991). Towards a paradigm shift in biology. Nature **349**: 99.

Jowett, T. (1986). Preparation of nucleic acids. pp. 275–286. In: Drosophila, *A Practical Approach*. D. B. Roberts, Ed. IRL Press, Oxford, UK.

Kessler, C., and V. Manta (1990). Specificity of restriction endonucleases and DNA modification methyltransferases—a review (Ed. 3). *Gene* **92**: 1–248.

Miller, J. H. (1992). *A Short Course in Bacterial Genetics. A Laboratory Manual and Handbook for* Escherichia coli *and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Neidhardt, F. C., Ed. (1999). Escherichia coli *and* Salmonella:*Cellular and Molecular Biology*. 2nd ed. Am. Soc. Microbiol., Washington, DC.

Pingoud, A., J. Alves, and R. Geiger (1993). Restriction enzymes. pp. 107–200. In: *Methods in Molecular Biology.* Vol. 16. *Enzymes of Molecular Biology*. M. M. Burrell, Ed. Humana Press, Totowa, NJ.

Sambrook, J., and D. W. Russell (2001). *Molecular Cloning. A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Snyder, L., and W. Champness (1997). *Molecular Genetics of Bacteria*. Am. Soc. Microbiol., Washington, DC.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503–517.

Southern, E. M. (2000). Blotting at 25. *Trends Biochem. Sci.* **25**: 585–588.

Some Relevant Web Sites

Many Web sites provide current protocols and other information. Some with links to many other sites include:

- AgBioTechNet (http://www.agbiotechnet.com)
- BioBencHelper (http://biobenchelper.hypermart.net)
- Information Systems for Biotechnology (http://www.isb.vt.edu)
- National Center for Biotechnology Information (http://www.ncbi.nem.nih.gov)
- REBASE 2000. Restriction Endonuclease Home Page (http://rebase.nab.com/cgl-bin/statist)

See also Web sites of major companies selling kits and reagents, as well as major institutes, societies and universities.

6

Cloning and Expression Vectors, Libraries, and Their Screening

6.1. Overview

Cloning of DNA has five essential components: 1) a method for generating exogenous DNA fragments, 2) reactions to join exogenous DNA fragments to a vector, 3) a method to introduce the vector into a host cell where the vector ensures the exogenous DNA is replicated, 4) methods for selecting or identifying the vectors that contain the introduced DNA (recombinant molecules), and 5) methods for analyzing the cloned DNA.

Genomic libraries can be constructed in λ phage and other vectors. Complementary DNA (cDNA) libraries can be cloned into various vectors, including cosmids or phagemids. Cloning into single-stranded M13 phage results in single-stranded DNA suitable for sequencing.

Cloning is feasible because a diverse array of enzymes are available to synthesize, ligate, and modify the ends of DNA molecules. Screening libraries for genes of interest can be accomplished by nucleic acid or antibody hybridizations, chromosome walking, and sequencing.

Cloning provides the basis for identifying specific genes, producing DNA copies of mRNA (= cDNA), and, in some cases, producing gene products (proteins) in *E. coli*, yeast, or insect cells by incorporating the DNA into expression vectors. Two insect baculoviruses have been genetically engineered to express proteins (expression vectors) and are used to produce proteins in insect cells in tissue culture or in intact lepidopteran larvae.

6.2. Introduction

The term "cloning" has multiple meanings. For example, Dolly the sheep is a clone because a nucleus of a donor somatic cell was inserted into an egg (from which the original nucleus was removed). That egg was implanted in a host ewe and produced a lamb (Dolly) that is genetically identical to the sheep donating the nucleus. In this chapter, **cloning** means that a single vector molecule (plasmid or phage or engineered versions of these) containing exogenous DNA is multiplied in cells so that multiple identical copies (clones) are produced. A **vector** is the agent used to replicate, or multiply, the exogenous DNA. Vectors are segments of DNA with an **origin of replication** so that the vector is replicated after it is introduced into a host cell. Vectors can be plasmids, bacteriophage, baculoviruses, or hybrid engineered molecules called cosmids and phagemids.

Chapter 5 introduced the use of plasmid vectors. This chapter introduces vectors derived from the *E. coli* bacteriophage λ, the single-stranded DNA bacteriophage M13, and engineered, hybrid vectors combining components from bacterial plasmids and λ called cosmids. Phagemids are engineered hybrid molecules that combine elements of plasmids and M13 vectors. BACs, bacterial artificial chromosomes, are used to clone very large segments of DNA.

The most commonly used host cell is *E. coli* but others are used, including the bacterium *Bacillus subtilis* or yeast *Saccharomyces cerevisiae*. Insect cells in tissue culture also can be used as hosts for baculovirus expression vectors (described below).

A multitude of vectors have been developed commercially for cloning. The choice of an appropriate vector depends upon the goal of your experiment. To develop a new vector requires an extensive knowledge of the biology and genetics of *E. coli* or other host cell, the plasmid or bacteriophage, and enzymology. It is impossible to describe all existing vectors and their uses; furthermore, such a description would become obsolete rather rapidly. New vectors are engineered and advertised regularly.

Basically, a vector is a segment of DNA with an origin of replication which allows it to be maintained stably after it is introduced into its host cell. Most vectors contain unique restriction sites in a region of the vector that contains nonessential genetic information. The unique restriction sites are where exogenous DNA fragments can be inserted into the vector. Often several cloning or restriction sites are combined into a single region, which is called a **polylinker** or multiple cloning site.

One widely used plasmid vector series is a derivative of pBR322 (plasmid BR322). It was described and its structure illustrated in Chapter 5 (Figure 5.2). Another plasmid vector series is the pUC group, which contains a functional segment of the *E. coli lacZ* gene (Figure 5.3). Thus, *E. coli* containing this plasmid produce blue colonies if provided with the appropriate substrate in the agar medium. If exogenous DNA is inserted into the cloning site, which is located within the *lacZ* gene, the *lacZ* gene sequence is disrupted, the gene product is no longer made, and the *E. coli* colonies are colorless. The pUC plasmids produce an increased copy number in *E. coli*, which results in an increased yield of recombinant DNA molecules compared to the pBR322 plasmid series.

Cloning can be used to produce gene libraries, develop mutated genes for experiments, provide single-stranded DNA for sequencing, and permit eukaryotic genes to be translated in *E. coli*, insect tissue culture cells, or lepidopteran larvae. Figure 6.1 identifies many

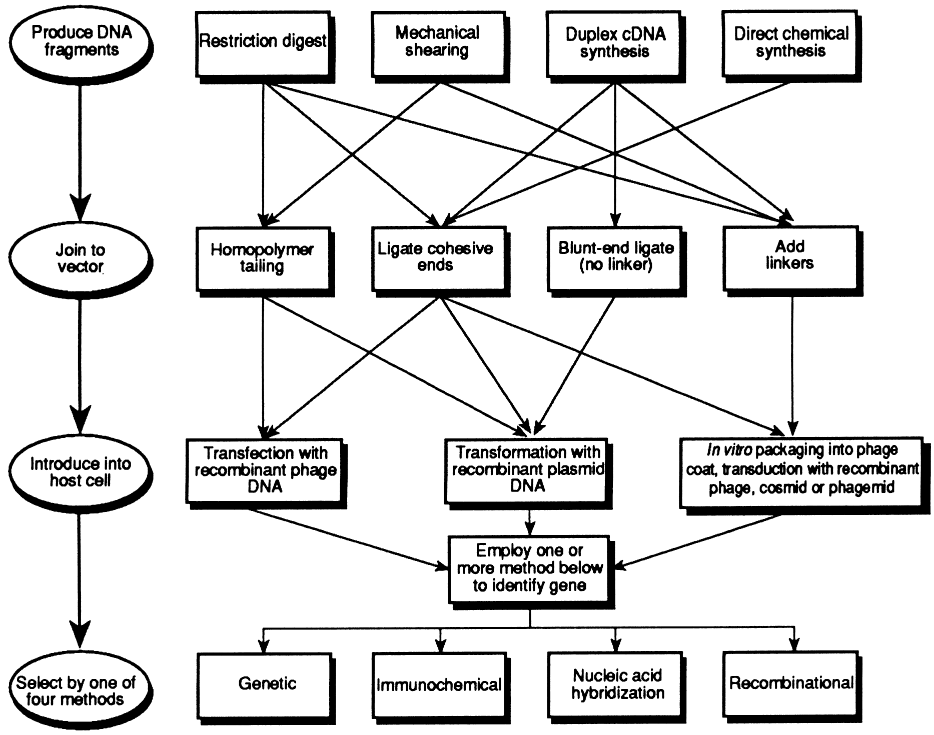


Figure 6.1. Generalized scheme outlining the steps employed in cloning DNA in *E. coli*. There are four major components: obtaining DNA fragments, joining them to the vector, introducing the recombinant molecule into an appropriate host cell, and identifying or selecting the recombinant DNA of interest. All of these steps can be achieved in several ways. Choosing an appropriate cloning scheme depends upon the goals of the experiment. (Revised from Old and Primrose 1989.)

of the steps and procedures involved in cloning, but a full description of all the techniques employed is beyond the scope of this chapter.

Commercial companies will provide genomic or complementary DNA (cDNA) libraries for a fee if you provide the DNA or RNA. The availability of the polymerase chain reaction (PCR, described in Chapter 8) makes the construction of libraries somewhat less important than formerly, especially if the goal is to isolate only one or a few genes. Complete libraries are essential if the entire genome is to be sequenced, however. References at the end of this chapter provide additional information and protocols on constructing libraries.

6.3. The Perfect Genomic Library

No library is perfect. However, the perfect **genomic library** would contain all the DNA sequences in the entire genome. The library would be stable and have a manageable number of overlapping cloned segments. The clones would contain sufficiently large DNA segments that they could contain whole genes and their flanking sequences. Ideally, the library could be amplified without loss or misrepresentation of sequences, and it could be stored for years without significant loss of information. Unfortunately, no single vector provides all of these desirable attributes.

Partially digested genomic DNA fragments can be cloned into bacteriophage λ relatively easily. The disadvantage is that the average λ library of an insect genome would contain >100,000 clones, each with an insert that averages ~15 to 20 kb. Cosmid vectors have an advantage in that the size of the inserted DNA can be two- to threefold larger, and therefore fewer cosmids need to be evaluated to find the gene(s) of interest. Thus, cosmids can provide a significant advantage when it is important to work with an entire gene and its flanking sequences.

Genomic libraries constructed from random fragments of DNA are called **"shotgun" libraries**. To generate a genomic library for an insect may require 10^6 clones to ensure that all sequences in the genome are represented. The formula for estimating the required size of the library is:

$$N = \frac{\ln(1 - P)}{\ln(1 - a/b)}$$

in which N is the number of clones required, P is the probability that a given sequence will be present, a is the average size of the DNA fragments inserted in the vector, and b is the total size of the genome.

The library can be read only if there is a key to open it. The key to libraries is some sort of probe, as is described later.

6.3.1. λ Phage as a Vector

Lambda (λ) is a genetically complex, but well-studied, double-stranded-DNA bacteriophage of *E. coli*. The entire DNA sequence of the λ chromosome has been determined. Nearly 40% of the 48.5-kb chromosome is not essential for propagating the phage in its host. If this nonessential DNA is removed, about the same amount of exogenous DNA can be inserted. At each end of the linear DNA molecule there are short, single-stranded 5′ projections of 12 nucleotides, called ***cos* sites**, that are complementary to each other in sequence.

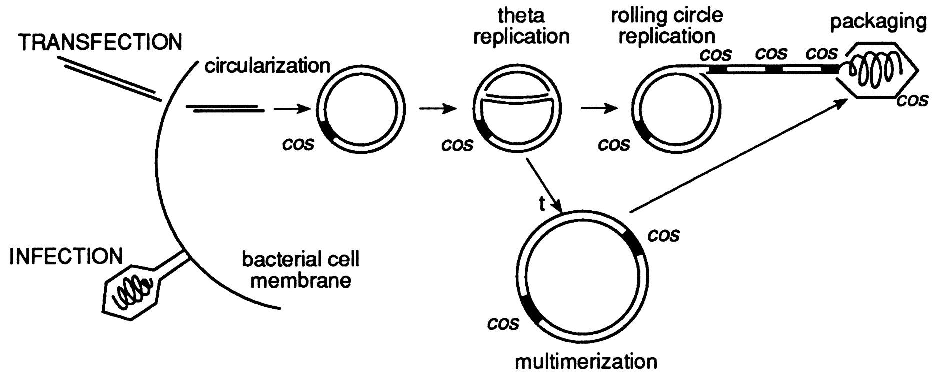


Figure 6.2. The typical lytic cycle of bacteriophage λ begins when the phage adsorbs to an *E. coli* cell by the tail fiber. The linear DNA molecule is injected into the host cell where it circularizes by annealing at the *cos* sites. The λ DNA replicates by a "rolling circle" mechanism, producing long concatenated molecules with the individual chromosomes annealed at the *cos* sites. Later in the infection cycle, phage genes for the protein head and tail are turned on. When these components are produced, the replicated DNA is packaged into the icosahedral head after the concatenated DNA is cut at the *cos* sites. If the DNA is too long or too short, it will not be packaged in the head, but will be lost. Lysis of the host cell then occurs, typically releasing approximately 100 progeny phage that enter new cells to replicate.

The *cos* sites enable the λ chromosome to circularize after the linear phage is injected into its *E. coli* host (Figure 6.2). After replication within the host cell, the λ DNA is in a linear form when it is packaged into a protein coat. The protein coat consists of an icosahedral head and a tail that ends in a tail fiber (Figure 6.2). The infective phage thus consists of the DNA molecule plus a protein head and tail. The protein coat allows a phage particle to adsorb by the tip of its tail fiber to receptor sites on the outer membrane of its host cell. Adsorption is temperature independent but dependent upon the presence of magnesium ions.

A wild-type λ has two phases to its life cycle. Although the temperate or **lysogenic** phase is of little interest to the genetic engineer, λ that has been genetically modified to serve as a vector retains many characteristics of the second, or lytic, phase.

In the **lytic** phase, early DNA transcription establishes the lytic process, middle genes replicate and recombine the DNA, and late genes produce protein for packaging the DNA into mature phage particles. Phage DNA is replicated in a "rolling circle" mode (Figure 6.2). Multiple copies of the replicated DNA molecules are assembled in a linear tandem array, with the termini of each molecule joined at the *cos* sites. The *cos* sites form the recognition site of a specific endonuclease which cuts the DNA during the packaging process so that a single DNA molecule is inserted into the head of the protein coat. In summary, in a lytic infection, the phage takes over the host cell machinery: phage DNA is replicated, head and tail proteins are made, the replicated DNA is packaged, and the host cell is lysed to release about 100 infective particles.

In the temperate or lysogenic phase, most phage functions are repressed and lysis is avoided. In lysogeny, the λ DNA is inserted into the host chromosome by site-specific recombination and the phage genome (called a prophage) is replicated as part of the *E. coli* chromosome. Nearly all λ vectors used in genetic engineering lack the ability to enter the lysogenic phase.

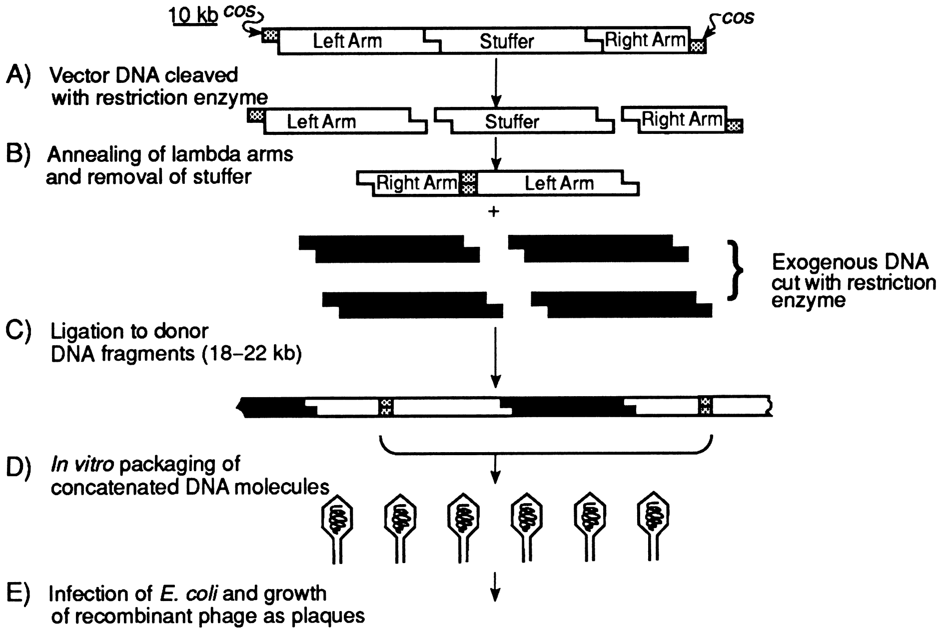


Figure 6.3. A schematic outline of a bacteriophage λ replacement vector. A linear molecule contains the *cos* sites, a left (L) and right (R) arm, and a "stuffer" region with nonessential DNA. The vector is digested with an appropriate restriction enzyme, the stuffer fragment is removed, and the two arms anneal. Exogenous DNA that has been cleaved with an appropriate restriction enzyme is added, and the fragment is ligated in. Exogenous DNA fragments 18 to 22 kb long can be incorporated because these molecules can be successfully packaged by *in vitro* packaging. *E. coli* is infected with the λ and thousands of individual plaques are produced. Each plaque contains many thousands of replicas (clones) of a single phage containing exogenous DNA.

λ has been genetically improved as a vector: 1) Genes in the central region of the chromosome that code for events associated with recombination and lysogeny have been deleted and replaced with exogenous DNA (Figure 6.3). 2) Vectors have been engineered to contain DNA sequences (cloning sites) to facilitate the insertion of exogenous DNA. Engineered versions are of two major types. **Insertion vectors** have a single target site at which foreign DNA can be inserted, while **replacement vectors** have a pair of sites defining a fragment that can be removed and replaced by foreign DNA (Figure 6.3).

Once exogenous DNA has been inserted into the vector λ, this molecule can be multiplied (cloned) by inserting it into host *E. coli* cells in one of two ways. Naked λ DNA (lacking a protein coat) can be introduced into *E. coli* cells in a process called transfection.

Transfection is the infection of bacteria by viral nucleic acid alone. The efficiency of transfection is $> 10^4$ recombinant clones per microgram (μg) of donor DNA. This efficiency would suffice for the construction of genomic libraries from species with small genomes. However, larger genomes, such as those of insects, require a more efficient method of inserting the vector DNA into *E. coli.*

The way to increase efficiency in introducing recombinant λ DNA molecules into *E. coli* is called ***in vitro* packaging**. By incorporating the recombinant DNA molecules into phage protein coats, *E. coli* can be infected much more readily, which increases the likelihood of producing complete genomic libraries. Efficiency of infection of *E. coli* with

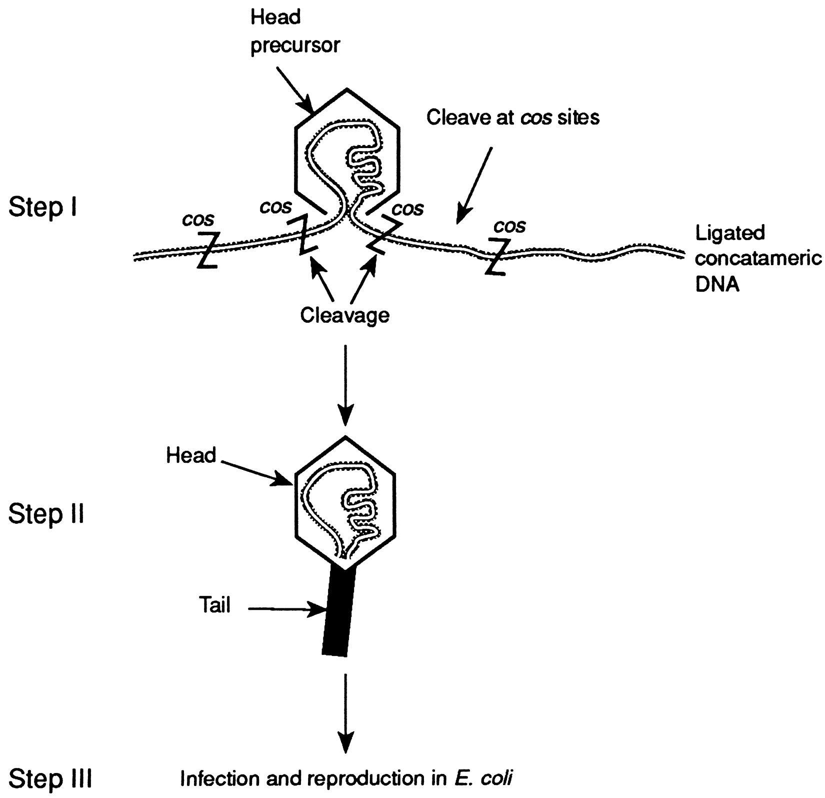


Figure 6.4. *In vitro* packaging of cloned DNA in λ involves providing a protein head and tail precursor. The DNA is cut at the *cos* sites, and if the DNA is ca 50 kb in length it will be packaged in the head. The complete phage is then used to infect *E. coli* and amplify the recombinant molecule.

packaged DNA can be 10^6 recombinants per μg of vector DNA, an increase in efficiency over transfection by nearly two orders of magnitude.

In vitro packaging involves a number of steps and specific conditions (Figure 6.4). One critical condition is the size of the inserted DNA. The amount of exogenous DNA inserted into the vector must be regulated carefully: the *cos* sites must be separated by DNA that is ~78 to 105% of the length of the wild-type chromosome. In an insertion vector, only 14 kb of DNA, or less, can be cloned. In a replacement vector, up to 22 kb of DNA can be inserted. In replacement vectors, a pair of restriction sites flank the nonessential central region of the phage DNA called the "stuffer region." When the stuffer region is excised and the insert DNA is ligated into this region, a DNA molecule is produced that can be packaged efficiently.

In vitro packaging requires the following components: 1) the DNA molecules to be packaged, 2) high concentrations of phage head precursor protein, 3) proteins that participate in the packaging process, and 4) phage tails. These packaging ingredients are obtained by combining a very concentrated mixture of the lysate from two different λ strains that are lysogenic. One mutant λ strain can progress no further in the packaging process than the pre-head stage because it carries a mutation in a gene (gene *D*) and therefore accumulates this precursor. The other mutant λ strain is prevented from forming any head structure by a mutation in a different gene (gene *E*), but can produce the tail component. In the



Figure 6.5. An agar plate with *E. coli* with plaques caused by the bacteriophage λ. Each clear area indicates where a single bacterium initially was infected with λ. After replication, the emerging λ attacked adjacent *E. coli* and lysed them, resulting in a clear "plaque" of killed *E. coli* and λ on the surface of the agar.

mixed lysate, both head and tail components become available so that a complete phage can be assembled that contains recombinant DNA.

Transfer of DNA into the *E. coli* host involves adsorption of phage to specific receptor sites on the outer membrane of the *E. coli*. Because phage will adsorb to dead cells and debris, only healthy bacterial cultures should be used to reduce loss of efficiency. Once the *E. coli* have taken up the phage, they are plated out on nutrient agar and allowed to grow at least overnight at 37°C.

Infected bacterial colonies grow, but clear areas (**plaques**) consisting of lysed cells will be seen surrounded by an opaque background of unlysed bacteria (Figure 6.5). Each plaque represents an original bacterial cell that was infected, ideally by only a single λ. Thus, each plaque should contain multiple copies of a *single kind* of recombinant DNA molecule. Even the smallest plaque is likely to contain sufficient phage DNA to be detectable by plaque hybridization, a probe technique described below.

A visual method for identifying plaques containing λ with recombinant DNA involves the use of the *lacZ* gene. This gene codes for part of the β-galactosidase enzyme, which cleaves lactose to produce glucose and galactose. Inserting exogenous DNA into this gene inactivates synthesis of β-galactosidase. To identify *E. coli* colonies containing recombinant phage, the agar is made up with a lactose analogue called **X-gal** (5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside). X-gal is cleaved by β-galactosidase into a product that is bright blue. If exogenous DNA has inserted into and disrupted the β-galactosidase gene, plaques appear white or colorless. Plaques containing λ without the exogenous DNA will produce the blue color.

Figure 6.6 outlines the steps involved in one strategy for producing a representative genomic library in a λ replacement vector. The genomic DNA and the vector DNA can be

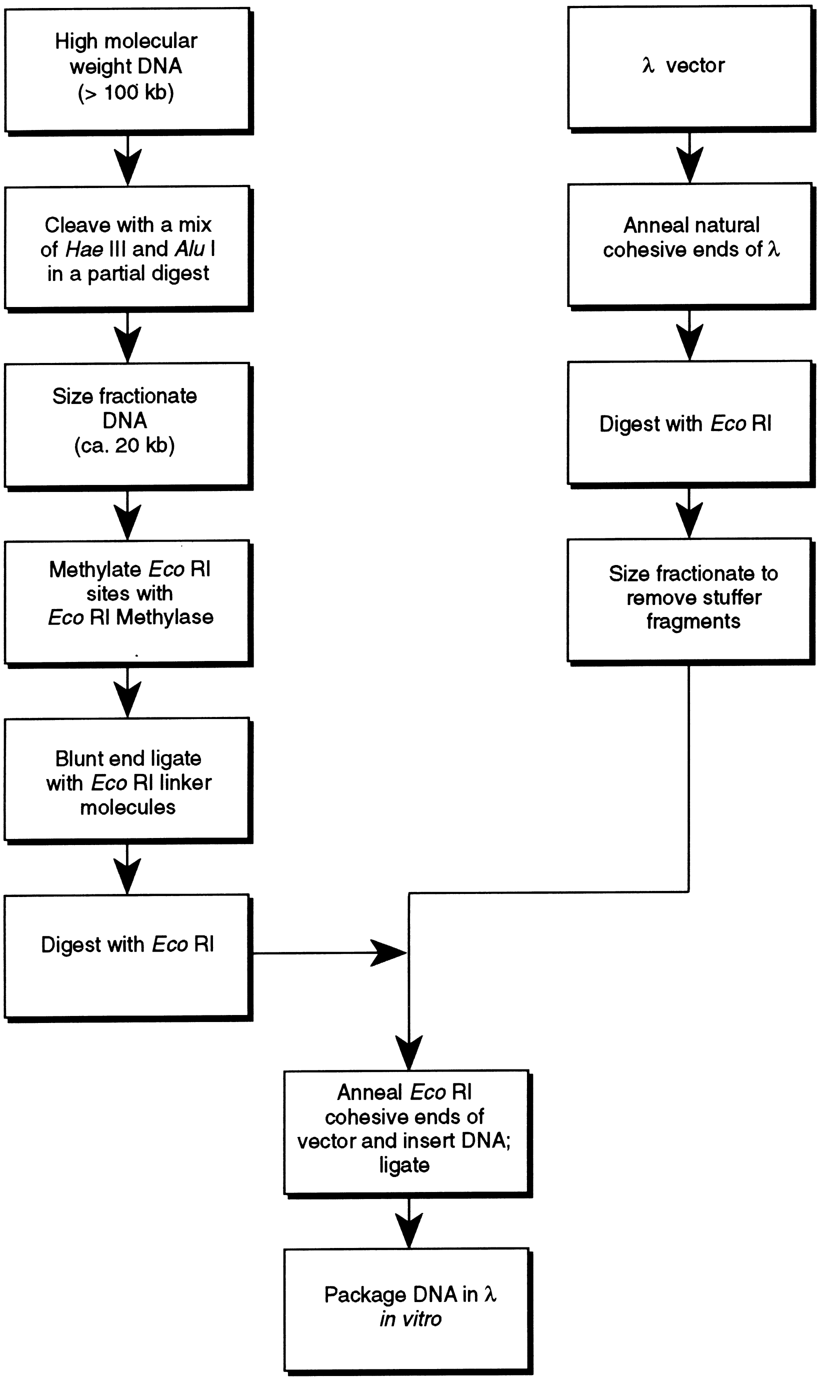


Figure 6.6. One method for producing a representative genomic library in a λ replacement vector. Two parallel processes are carried out: preparation of the exogenous DNA and preparation of the vector DNA. These are then ligated together and packaged *in vitro*. The specific restriction enzymes and ligation method can be varied.

prepared simultaneously. In this example, the genomic DNA is cut with a mixture of two restriction endonucleases (*Hae*III and *Alu*I) in such a manner that the DNA is only partially digested. The genomic DNA is then sized so that fragments about 20 kb long are isolated. Meanwhile, the λ vector DNA is digested with the restriction enzyme *Eco*RI and purified so that the stuffer fragments are removed. The genomic DNA has linker molecules added to it prior to the annealing reaction. When the genomic DNA and the vector DNA are combined, they anneal at their complementary cohesive ends and are ligated together. The last step involves providing a protein coat for the DNA by *in vitro* packaging.

Commercial cloning kits simplify the procedures considerably because such kits provide vectors, enzymes, *in vitro* packaging materials, and detailed protocols. It is even simpler to supply genomic DNA to a company that will provide a complete genomic library for a fee. However, a few points should be made about constructing a genomic library.

The genomic DNA to be cloned must be of high molecular weight and not excessively sheared during its isolation from the insect. High-molecular-weight DNA is needed because the DNA will be partially digested with a restriction enzyme to generate a random collection of DNA fragments and these need to be at least 20 kb long. The DNA to be digested actually must be longer than 20 kb so that after digestion both ends of the fragment will have cohesive ends. DNA fragments with only one cohesive end (and one broken end) cannot be inserted into vectors. DNA shorter than 20 kb won't be packaged into the phage and will be lost. Thus, DNA extraction should be carefully executed to avoid damaging or shearing the DNA.

The genomic DNA fragments ideally will be representative of the entire insect genome. If the restriction enzymes that are used cut relatively frequently compared with the desired fragment size, a *partial* digestion will produce a set of overlapping fragments. Ideally these fragments will be a nearly random array of the entire genome. However, it is possible that some regions of the genome will not be represented because they lack the appropriate cleavage sites for the enzyme used, or the DNA may not be cleaved with equal efficiency, particularly in heterochromatic regions. Furthermore, some regions of the genome may be toxic to their *E. coli* host cells.

After the genomic DNA has been partially digested with an appropriate restriction endonuclease, the DNA fragments are size-fractionated by centrifugation or by gel electrophoresis. This separates out the DNA fragments greater than or smaller than 18 to 22 kb. Preparing a representative library requires high-quality vector DNA. Large-scale preparation of λ DNA should yield pure preparations that lack the central stuffer region, or it can reinsert back into the vector later on. Removal of the central stuffer region is carried out by centrifugation, elution, or electrophoretic separation. However, it is difficult to remove all of the stuffer fragment, so it is important to determine, by appropriate controls, how often the stuffer is reinserted back into the vector.

The ligation reaction must be carefully regulated by optimizing concentrations of vector and exogenous DNA. Because a portion of the DNA molecules to be ligated will have damaged ends, the ratio of vector to insert DNA will probably have to be determined empirically in small reactions. It is desirable to produce long concatenated molecules that can be cut at the *cos* sites and packaged. Likewise, the appropriate ratios of ingredients used for *in vitro* packaging will have to be determined empirically. Once the DNA has been packaged, the phage can be stored at 4°C for years. Alternatively, the phage can be amplified by multiplication in *E. coli*.

Commonly used vectors derived from λ include the gt and EMBL series. λ gt10 was designed for cloning short DNA fragments, especially cDNA. λ gt11 is used to construct cDNA libraries, as described below. DNA properly aligned with the *lacZ* gene in λ gt11

will be expressed in *E. coli* as a **fusion protein**. EMBL vectors are a family of replacement vectors that provide a high level of reproduction in *E. coli*, polylinker cloning sites, and the ability to select for recombinant phage. EMBL3 and EMBL4 vectors, or their derivatives, are particularly useful for constructing genomic libraries.

6.3.2. Cloning with Cosmids

Cosmids are engineered vectors that combine characteristics of both plasmids and phage. They have been constructed to include a fragment of λ that includes the *cos* site (Figure 6.7).

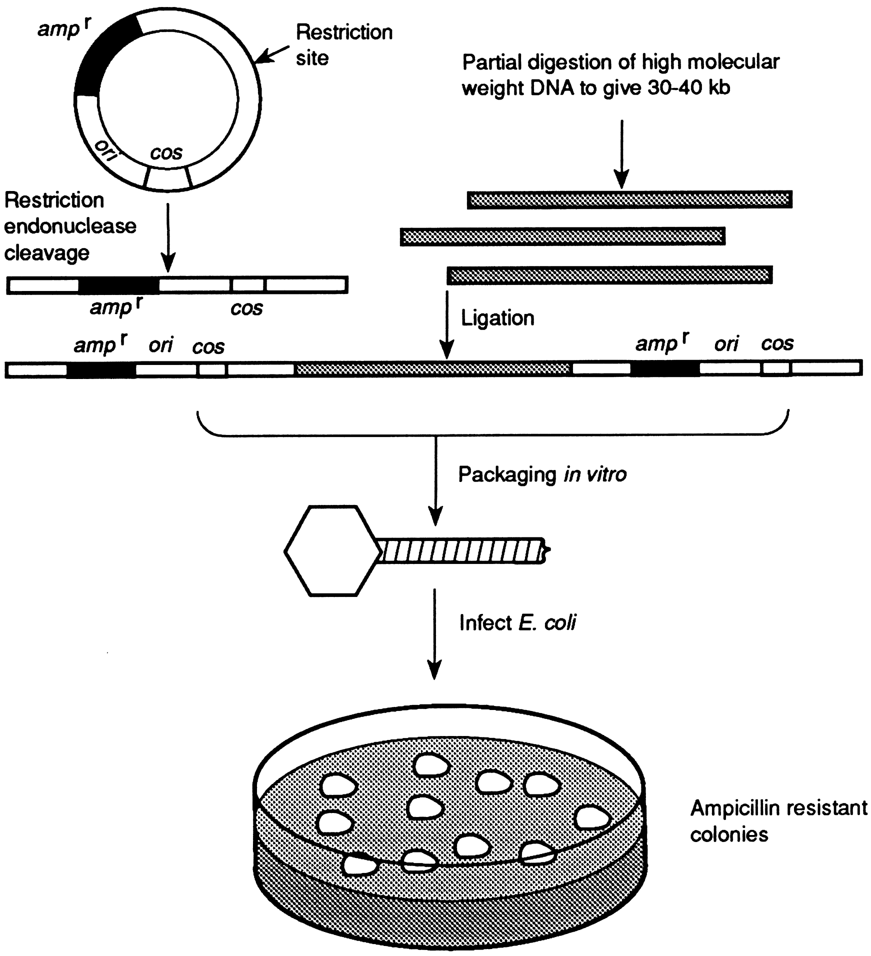


Figure 6.7. Outline of procedures used in cloning with a cosmid vector. This vector contains a *cos* site, a restriction site for inserting exogenous DNA, and a gene for ampicillin resistance. Exogenous DNA is cut with an appropriate restriction enzyme, as is the vector. The vector and exogenous DNA are ligated together, producing a recombinant molecule 37–52 kb long, which can be packaged in λ by *in vitro* packaging. The packaged vector infects *E. coli*, injecting its DNA into the host, where it circularizes and multiplies. *E. coli* cells that receive the cosmid are distinguished from cells that are not infected by their ability to survive on media containing ampicillin.

Cloning into cosmids is similar to cloning in λ. It involves digesting exogenous DNA with a restriction enzyme, cutting the cosmid vector with a compatible restriction enzyme, combining the two, and ligating them.

Once the exogenous DNA is inserted into the cosmids, cosmids are packaged in a manner similar to that employed with λ. Packaging the cosmid recombinants into phage coats provides a useful method for selecting the size of the inserted DNA. What is significant about cloning with cosmids is that larger DNA fragments, 32 to 47 kb, of foreign DNA can be inserted into the vector and still be packaged.

After *in vitro* packaging, cosmids are used to infect a suitable *E. coli* strain. Infection of *E. coli* involves injection and circularization of the cosmid DNA, but no phage protein is produced. Transformed *E. coli* cells are identified on the basis of their resistance to a specific antibiotic.

While having a large capacity for DNA fragments is a benefit in cloning with cosmids, it can be a detriment. If, during a partial digestion with restriction enzymes, two or more genomic DNA fragments join together in the ligation reaction, a clone could be created with fragments that were not initially adjacent to each other. This can be a problem if the researcher is interested in the relationship between a gene of interest and the DNA surrounding it. The problem can be overcome by size fractionating the partial digest. However, even then, cosmid clones could be produced that contain noncontiguous DNA fragments ligated to form a single insert. This problem can be solved by dephosphorylating the foreign DNA fragments to prevent them from ligating together, but this makes cosmid cloning very sensitive to the exact ratio of insert and vector DNAs. If the ratio is unbalanced, vector DNAs could ligate together without containing any exogenous DNA insert. This is resolved by treating the vector to two separate digestions, which generate vector ends that are incapable of ligating to each other after phosphatasing.

Commonly used cosmid vectors include the pJB8 and the pcosEMBL family. The pcosEMBL family was designed to simplify isolation of specific recombinants from cosmid libraries and speed up isolating large regions of complex genomes in an ordered array of overlapping clones (i.e., chromosome walking, which is described below). The vectors in this family differ by having different cloning sites and different numbers of *cos* sites. In our theoretical experiment, recombinant cosmids can be identified by rearing *E. coli* in the presence of the antibiotic ampicillin (Figure 6.7).

6.3.3. Cloning in the Filamentous Phage M13

M13 is a filamentous phage of *E. coli* that contains a circular single-stranded (ss) DNA molecule that is 6407 nucleotides long. M13 infects only strains of *E. coli* that have **F pili** because the site where these phage adsorb appears to be at the end of the F pilus (Figure 6.8). Replication of M13 does not result in host-cell lysis. However, the infected cells grow and divide more slowly and extrude up to 1000 virus particles into the medium.

Replication of M13 phage (which is single-stranded) involves conversion of the DNA to a double-stranded (ds) or replicative form (RF). The ds RF multiplies until about 100 RF molecules are produced within the cell (Figure 6.8). The replication of the RF then becomes asymmetric because of the accumulation of a viral-encoded binding protein that is specific to ss DNA. The binding protein binds to the M13 DNA and prevents synthesis of a complementary strand. Subsequently, only ss viral DNA is synthesized, and these are extruded from the host cell. As the ss M13 DNA molecules move through the *E. coli* cell membrane, the DNA binding protein is removed and the M13 DNA is coated with capsid protein.

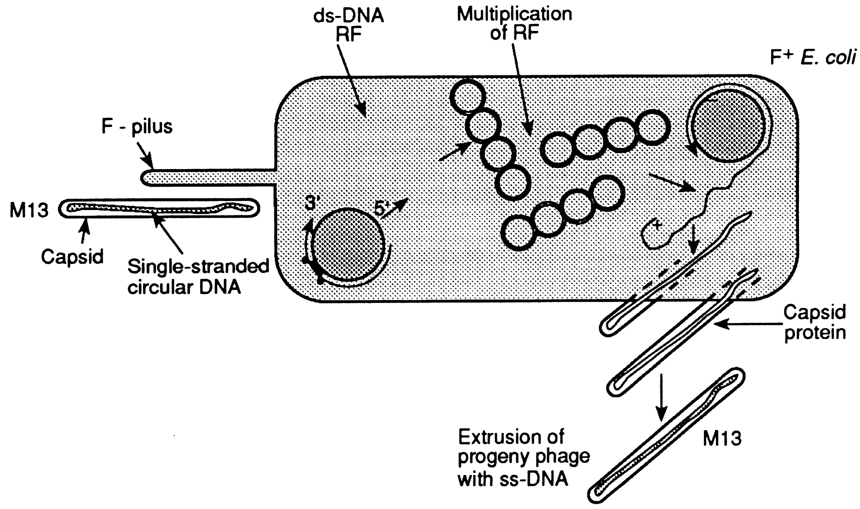


Figure 6.8. An outline of the infection cycle of the bacteriophage M13. The single-stranded phage attaches to the F-pilus of *E. coli,* injects its DNA into the host, and begins to produce approximately 100 copies of double-stranded (RF) molecules. DNA replication then shifts to producing ss DNA molecules, which are extruded through the host cell wall, during which time they are coated with a capsid protein coat. M13 has been engineered as a vector and is used to produce ss DNA molecules, which is particularly useful in DNA sequencing reactions.

M13 has a number of advantages as a vector. First, ss DNA is required in a number of applications, including the dideoxy DNA sequencing method (described in Chapter 7). Second, ss M13 vectors allow the genetic engineer to combine cloning, amplification, and strand separation of a ds DNA fragment in one operation. Third, because the phage DNA is replicated in a ds circular (RF) intermediate stage, the RF DNA can be purified and manipulated just like a plasmid. Fourth, both RF and ss DNA will transfect competent *E. coli* cells and yield either plaques or infected colonies. Fifth, it is possible to package DNA up to six times the length of the wild-type M13 DNA.

The M13 phage does not contain excess DNA that can be removed. However, there is a 507-bp region which contains the origins of replication for the viral DNA and its complementary strand. In most vectors derived from M13, foreign DNA has been inserted at this site. M13 vectors also have been modified by inserting polylinkers and the *lacZ* gene so that white plaques are formed instead of blue if exogenous DNA has been inserted into the *lacZ* cloning site.

6.3.4. Phagemids

A phagemid is an engineered vector that contains plasmid and M13 components. Phagemids provide another method for obtaining ss DNA. Phagemids carry two replication origins, one a standard plasmid origin and the other derived from M13. The M13 origin is crucial for the synthesis of ss DNA. However, production of ss DNA requires enzymes and coat proteins coded by phage genes, which are lacking in phagemids. As a result, cells containing a phagemid vector must be coinfected with a helper phage if ss DNA is desired. The helper phage converts the phagemids into ss DNA molecules, which are then assembled into phage particles and secreted from the cell.

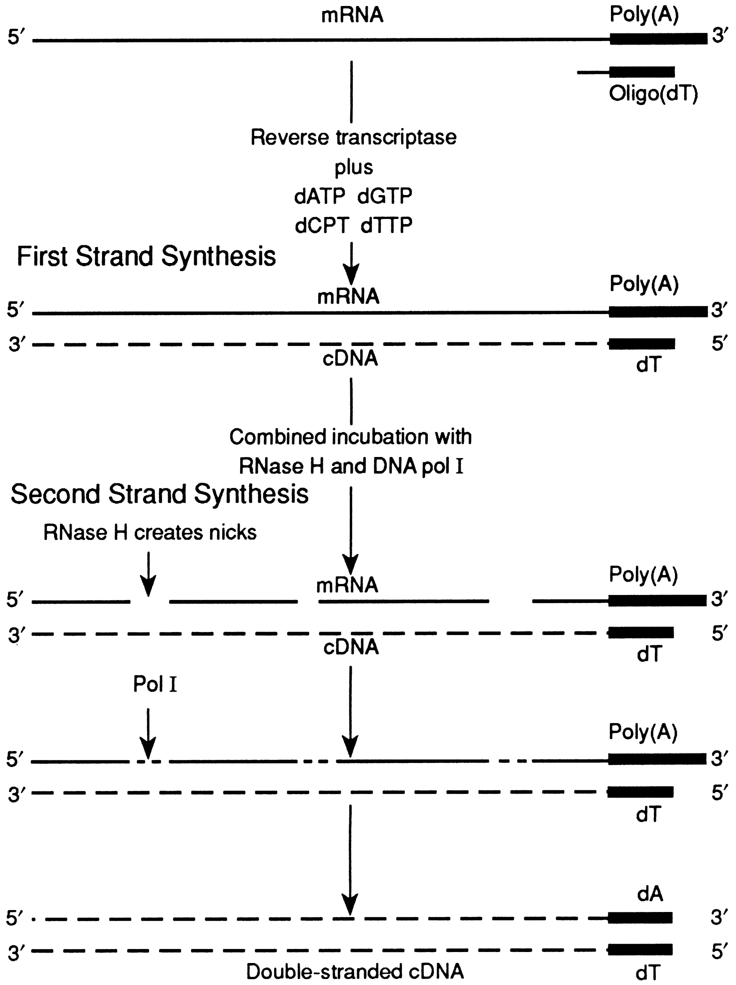


Figure 6.9. cDNA cloning involves two steps. In the first step an oligo(dT) anneals to the poly(A) region of mRNA. The enzyme reverse transcriptase and dATPs, dCTPs, dGTPs, and dTTPs are provided to produce the first cDNA strand. Synthesis of the second strand of the cDNA involves additional incubation with RNase H and DNA polymerase I. The double-stranded cDNA produced is a complement to the mRNA, and thus lacks introns or regulatory sequences.

Many of the newer vectors are phagemids. Figure 6.9 illustrates one phagemid, pUC118. Foreign DNA can be cloned into pUC118 and propagated as ds DNA. There is a 476-bp fragment of M13 in these vectors, as well as a gene for ampicillin resistance. If cells carrying the phagemid are infected with the helper virus M13K07, then phage particles are produced containing ss DNA.

6.3.5. BACs

Bacterial artificial chromosomes (**BACs**) were developed so that scientists could insert very large segments of DNA and have it cloned in *E. coli*. BACs can incorporate an average of 140 kb of exogenous DNA and are based on the *E. coli* F-factor plasmid, which maintains a

low copy number in bacterial cells, thus minimizing the possibility of recombination which could result in the production of chimeric clones (Shizuya et al. 1992). Basic BACs do not contain a selection system or reporter genes suitable for expression in eukaryotic cell lines. However, BACs have been modified to be used as expression vectors in eukaryotic cells (Kim et al. 1998).

BACs have been extensively used in the Human Genome Project. The advantage of the BAC vector is that it can maintain up to 300 kb of DNA, and the clones are highly stable in their host, even after 100 generations. BACs thus facilitate the construction of DNA libraries of complex genomes because they allow fuller representation of all sequences.

6.4. cDNA Cloning

cDNA is DNA that is *complementary* to the mRNA that is transcribed from the gene. Because the mRNA is processed, cDNA will lack the introns usually found in insect genes. Thus, a cDNA clone will contain the DNA sequence of the protein of interest, but will lack introns and probably not contain the control sequences that regulate gene expression. cDNA cloning is used to produce a cDNA library or to produce probes for screening genomic libraries.

A **cDNA library** allows the genetic engineer to clone *only those genes that are active at a specific time or in specific tissues*. Genes that are not actively transcribed into mRNA will not be represented in the cDNA library. Thus, a cDNA library usually contains fewer clones than a genomic library. The gene of interest may occur in a frequency of 1×10^3 or 1×10^4. By contrast, a single-copy gene may be present in a genomic library in a frequency of only 1×10^5 to 1×10^6. Another benefit of a cDNA library is that it is possible, if an appropriate **expression vector** is used, to express a gene in a host such as *E. coli* or yeast. This enables the genetic engineer to produce large amounts of a specific gene product.

The quality of a cDNA library depends on the quality of the mRNA used as the template and the fidelity with which it can be **reverse transcribed** into cDNA. Messenger RNA, together with a suitable primer, and a supply of deoxyribonucleoside triphosphates can be converted into a ds DNA molecule with the enzyme **reverse transcriptase** (Figure 6.9).

The cDNA cloning process involves two steps: 1) the first strand of cDNA is produced, and 2) a strand that is complementary to the first strand is synthesized, so that a ds DNA molecule is produced. The primer used to synthesize the first DNA strand is usually an oligonucleotide consisting of deoxythymidine (dT) because it can hybridize to the $3'$ poly(A) tails of template mRNA and thus give rise to full-length copies of ds DNA. Once the ds cDNA molecule has been synthesized, it can then be inserted into a plasmid or phage vector that is capable of replicating in *E. coli* and, in some cases, of being translated into a protein.

A key to producing cDNA is the enzyme **reverse transcriptase**. Reverse transcriptase is capable of two functions *in vitro*: a polymerase activity and a ribonuclease H activity. The polymerase activity requires: 1) a template RNA molecule hybridized to a DNA primer with a $3'$-OH group, and 2) all four dNTPs to synthesize a DNA molecule which is a faithful complement of the RNA.

Cloning a cDNA library is more complex than cloning a genomic library into λ or cosmids. Before beginning the process, the goals of the project must be carefully considered and the basic approach chosen after deciding how the cDNA library will be screened to identify the gene(s) of interest (Kimmel and Berger 1987). For example, if antibodies will be used to identify clones capable of synthesizing specific peptides, the cDNA should be cloned into expression vectors.

The cloning techniques vary in the type of primer used, the method for second-strand synthesis, and methods for coupling the cDNA to the vector, which can be either a plasmid or λ. Commercially available reverse transcriptases can synthesize copies of mRNA sequences that are more than 3 kb long. However, the transcripts often terminate prematurely, making clones containing the 5′-end of the mRNA rare.

Figure 6.9 outlines the synthesis of double-stranded cDNA from mRNA. Messenger RNA is often prepared for cDNA cloning by affinity chromatography on oligo(dT) cellulose. The reaction is preceded by a brief heat denaturation of the mRNA to eliminate its secondary structure, because reverse transcriptase is inhibited if the mRNA exhibits a secondary structure. The polyadenylated mRNA, the primer, and the reverse transcriptase are combined. The primer in this case is a short sequence of (dT) residues. The product of the first-strand synthesis is a hybrid of mRNA and the synthesized cDNA. The first strand is used as a template–primer complex to make the second strand of DNA. The enzyme RNase H is used to introduce gaps in the mRNA strand. At the same time, DNA polymerase I uses the primer–template complexes formed by RNase H to synthesize a double-stranded DNA.

Once the double-stranded DNA is synthesized, it is inserted into a vector. To insert it into a vector, the synthesized molecule needs to have ends that can be ligated into the vector. One option is to make the cDNA blunt-ended by end filling with the Klenow fragment of DNA polymerase I and then ligating it into a vector that has been cut with a restriction enzyme that produces a blunt end. Another option involves the addition of cohesive ends to the cDNA so that it will ligate into the vector more easily. There are three methods to add cohesive ends to the ds DNA: 1) tailing with terminal transferase, 2) adding linkers, and 3) adding adaptors. The details of carrying out these procedures can be found in many cloning protocols.

There are about 10,000 *different* mRNA molecules in an average insect cell. At least 200,000 cDNA clones should be generated to be sure that a representative cDNA library is constructed. If the desired clone is a single-copy gene, then it will be rare, so powerful screening methods are required to isolate the clone of interest.

Isolating RNA is more difficult than isolating DNA. Preparation of mRNA requires the absolute elimination of **ribonucleases** (RNases) from glassware, pipettes, tips, and solutions. Anything that might contaminate the reactions with RNase must be eliminated, including hair, dust, and sneezes. Even fingerprints contain enough RNase to degrade your RNA. Furthermore, RNase is a very hardy enzyme and difficult to eliminate. Phenol extraction followed by ethanol precipitation was a common technique for isolating RNA, but various kits designed for that purpose are available now. Once RNA has been isolated, it must be evaluated for quality, often by agarose gel electrophoresis.

6.5. Enzymes Used in Cloning

A number of enzymes used in genetic engineering have been mentioned in this and previous chapters. Table 6.1 summarizes their name, principal activity(ies), source, and functions in genetic manipulations. Enzymes used to synthesize DNA include terminal transferase, DNA polymerase I, and reverse transcriptases. Enzymes that modify DNA include S1 nuclease, exonuclease III, Bal31 nuclease, and DNase I. There are more than 1400 restriction endonucleases that can cleave DNA in a predictable manner. T4 and *E. coli* DNA ligases join DNA molecules. Calf intestinal phosphatase (CIP) and T4 polynucleotide kinase are used to modify the 5′ ends of DNA molecules.

Table 6.1. Enzymes Useful for DNA Manipulation

Enzyme type Name (Source)	Functions in genetic engineering
Enzymes that synthesize DNA	
DNA polymerase I (*E. coli*)	5′ to 3′ DNA synthesis of template DNA with a primer; exonuclease functions (5′ to 3′ & 3′ to 5′) used to correct errors in DNA synthesis *in vivo*; generate labeled DNA probes by nick translation; synthesize cDNA
Klenow fragment of DNA polymerase I (*E. coli*)	DNA synthesis without 5′ to 3′ exonuclease ability; makes ds DNA from ss DNA; used in dideoxy sequencing; DNA labeling by random priming or end filling; converts 5′-overhangs of DNA cut with restriction enzymes to blunt ends
T4 DNA polymerase (phage T4)	Exonuclease in 3′ to 5′ direction; fill in overhanging ends of DNA cut with restriction enzymes
T7 DNA polymerase (phage T7)	3′ to 5′ exonuclease activity used in DNA end-labeling; converts 3′ overhangs to blunt ends
Taq DNA polymerase (*Thermus aquaticus*)	DNA synthesis at 60–70°C in PCR; several cloned versions are available, as are DNA polymerases from other microorganisms
Pfu DNA polymerase (*Pyrococcus furiosis*)	This polymerase and others have 3′ to 5′ exonuclease activity, which allows them to remove mismatches; used to amplify DNA fragments up to 40 kb by the PCR
Reverse transcriptases (from several RNA tumor viruses)	Synthesize copies of DNA from ss mRNA or DNA with template; cDNA synthesis most important
Terminal transferase (mammalian thymus)	Adds residues to any free 3′-terminus; used to add poly(dG) and (dC) to two DNA molecules to be joined
Enzymes that degrade DNA	
S1 nuclease (*Aspergillus*)	Degrades ss DNA endonucleolytically; removes projecting 3′ regions of ss DNA in cloning and S1 mapping
Exonuclease III (*E. coli*) Bal 31 nuclease (*Alteromonas espejiana*)	Degrades 1 of 2 strands of ds DNA from 3′ end of a blunt-ended double helix or from a projecting 5′ end; degrades both strands of ds DNA with blunt ends
DNase I (pancreas)	Introduces random nicks in ds DNA prior to labeling by nick translation; produces random fragments for shotgun cloning and sequencing in M13; study chromatin structure; study DNA–protein complexes
Enzymes that join DNA	
T4 DNA ligase	Seals ss nicks between adjacent nucleotides in ds DNA molecule, requires ATP; used to ligate two restriction fragments of DNA together in cloning
E. coli DNA ligase	Ditto, but requires NAD^+
Enzymes that modify the 5′ ends of DNA	
Calf intestinal phosphatase	CIP removes 5′-phosphate groups to generate an OH-terminus; prevents unwanted ligation of DNA fragments during cloning; used for end-labeling DNA probes
T4 polynucleotide kinase	Adds phosphates to 5′-OH ends; used in chemical cleavage method of DNA sequencing; used to add linkers or adapters in cloning
Enzymes that cut DNA	
Restriction endonucleases (many bacteria)	Type I, II, and III, >1400 types known; cleaves DNA; produces predictable termini, either blunt, 5′-overhang, or 3′-overhang

6.6. Isolating a Specific Gene from a Library

The production of a library is only a first step. The information in a library can be obtained only if the library can be screened. Screening identifies specific genes and provides information about genome organization, or about gene regulation. The ability to screen a library is dependent upon the availability of a **probe**. As pointed out in Chapter 5, a probe often is a molecule labeled with radioactive isotopes.

There are four ways to obtain a suitable probe for your library: 1) The amino acid sequence of the protein is known for the species being studied, or a related species, and can be used to predict and synthesize the sequence of an oligonucleotide hybridization probe. Because the genetic code is degenerate, the probe used may actually incorporate a mixture of oligos with optional bases, especially in the third site of the codon. 2) The gene of interest has already been cloned from a related organism, so that it can be used as a heterologous hybridization probe. For some genes, particularly the housekeeping genes, conservation of functional domains in proteins has been extensive, so that probes from other species can be used effectively. 3) The protein is abundant in a particular tissue so the relevant clone can be identified by its relative abundance in a tissue-specific cDNA library. 4) If the protein has been purified, it can be used to generate an antibody against it. The antibody can be used to identify recombinant cells for the presence of the specific enzymes.

Once a probe is obtained and labeled, it is used in DNA hybridization experiments to identify those clones that contain the DNA of interest. **Hybridization** involves immobilizing DNA samples from different clones on a solid support (such as a nitrocellulose or nylon membrane) and then probing the unknown DNA with a DNA or RNA sequence to identify the clones that contain the sequence of interest. Identification is possible because the radioactively labeled probe can base-pair with the desired DNA and then be detected by autoradiography. There are a number of different DNA hybridization techniques, including the Southern blot analysis described in Chapter 5. Another is plaque screening.

Figure 6.10 illustrates **plaque screening** of *E. coli*. First, *E. coli* are infected with λ that contain exogenous DNA and allowed to grow on an agar substrate. A nitrocellulose filter is laid onto the *E. coli* lawn and plaques. The precise orientation of the filter is marked. Some of the phage in the plaques become adsorbed onto the filter, where they release their DNA. The DNA is denatured by an alkali treatment and then brought to a neutral pH. After denaturation, the now single-stranded DNA on the filter is incubated with a radiolabeled probe. The probe base-pairs with the specific nucleotide sequence from the gene of interest, but not with DNA from plaques containing other genes.

The position of the probe that is hybridized to the immobilized DNA on the filter is located by autoradiography. The filter and the original agar substrate are then compared using the marker, and the corresponding plaque is located on the original agar substrate. A few phage can be picked from each plaque that yielded a spot on the X-ray film. The phage from that plaque are used to infect individual new *E. coli* cultures to produce multiple copies of that phage.

Plaque hybridization allows several hundred thousand plaques to be screened at once, and so a single-copy gene can be isolated from thousands of clones. Because the DNA that was inserted into the λ vector was cut at random, it is likely that more than one clone (plaque) will contain the DNA of interest. Ideally, at least one clone isolated by the probe will contain the complete gene, but this can only be determined after the DNA has been sequenced, a technique that is described in Chapter 7.

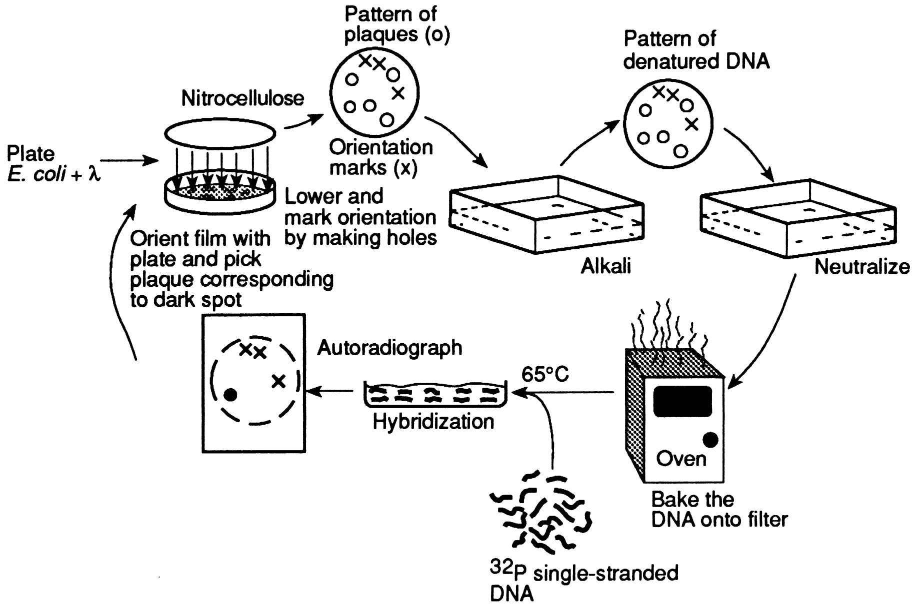


Figure 6.10. Plaque screening can locate specific genes. Plaque screening involves *in situ* hybridization of *E. coli* that have been transformed with a λ vector. *E. coli* infected with recombinant phage are plated out. A nitrocellulose filter is laid on top of the bacterial lawn and plaques. Some of the phage in a plaque adsorb to the filter. The filter is treated with an alkali to denature the phage DNA, neutralized, baked in an oven to immobilize the DNA, and placed in a solution with a radiolabeled DNA probe. The probe base pairs with sequences in the DNA that are complementary and identifies plaques that may contain the gene of interest. It is then possible to go back to the plate, pick a few phage from those plaques, and multiply them in *E. coli.*

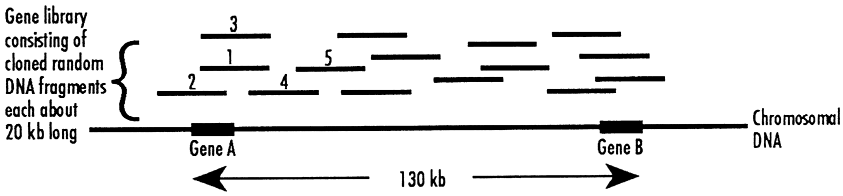


Figure 6.11. Chromosome walking is used to identify a gene of interest when a probe is not available. It can only be carried out when it is known that the target gene is linked to another gene which has been cloned and sequenced. First a clone containing gene A is isolated in fragment 1. This fragment is sequenced, and new probes are synthesized that contain sequences from each end of the fragment. The new probes are used to identify overlapping DNA clones in the library on each side of fragment 1, i.e., clones 2, 3, and 4. Clone 4 can, in turn, be sequenced, and used as a probe to identify clone 5, and so on until gene B is reached.

Another technique employed to identify specific DNA sequences is called **chromosome walking** (Figure 6.11). It is particularly useful with *Drosophila* but less useful with other insects for which less genetic information is available. Chromosome walking is used to isolate a gene of interest for which no probe is available. The gene of interest *must* be linked to a marker gene that has been identified and cloned. This marker gene is used as a probe to screen a genomic library. All fragments containing the marker gene are selected

and sequenced. The fragments are then aligned, and those cloned segments farthest from the marker gene in both directions are subcloned for the next step. The subclones are used as probes to screen the genomic library again to identify new clones containing DNA with overlapping sequences. As the process is repeated, the nucleotide sequences of areas farther and farther away from the marker gene are identified, and eventually the gene of interest will be found. As shown in Figure 6.11, the goal is to identify gene B, for which no probe is available. However, sequences of a nearby gene (A) are available in cloned fragment 1. In a large and random genomic DNA library, many overlapping cloned fragments will be present. Thus, clone 1 can be used as a probe to identify overlapping clones 2, 3, and 4. Clone 4 subsequently can be used as a probe to identify clone 5 until gene B is reached.

Once a gene has been identified in a genomic library, its DNA sequence can be determined. However, a DNA sequence by itself is of limited value. If you do not know something about the gene product, it may be difficult to determine unambiguously which sequences are the coding regions and which are introns. Intron boundaries may be established based on similarities to sequences of known introns (consensus sequences). If the gene product is unknown, it may be possible to identify the sequenced gene's function by comparing the DNA sequence with other sequences in DNA databases, although a fully convincing match is not always found. Thus, going from clone to DNA sequence to gene product may be a challenge, particularly if genes are being studied for which there are no known gene products.

One solution is to attempt to express the gene in order to obtain a gene product. To be expressed, genes require a promoter and, often, upstream control sequences. A variety of expression vectors have been developed to express cloned genes in *E. coli*. Such vectors require *E. coli* promoters if the eukaryote sequence is to be expressed. A detailed description of *E. coli* expression vectors is provided by Pouwels (1991). Of particular interest to entomologists is the use of baculovirus expression vectors to express insect genes, as well as other eukaryotic genes, as described below.

cDNA libraries can be screened by hybridization screening or by expression screening. If a protein of interest has been purified and part or all of the protein sequence has been obtained, then it is possible to predict the sequence of synthetic oligonucleotides that can be used as a hybridization probe to detect the appropriate cDNA clone(s). Alternatively, if an antibody to the protein is available, it can be used to identify the clone(s) of interest if the cDNA library is cloned into an expression vector (Figure 6.12). This vector contains the *lacZ* gene of *E. coli* and has a unique *Eco*RI restriction site near the end of the gene. cDNA cloned into this site in the correct orientation and reading frame will produce a fusion protein. Upon lysis of the *E. coli* cells, the protein is released and picked up on nitrocellulose in just the same way as in plaque screening. The plaque containing the interesting cDNA clone can be detected by incubating the filter with a specific antibody.

It is also possible to determine the difference in abundance between two different mRNA populations. Thus, mRNA produced from different tissues from the same organism or mRNA produced from a tissue before and after a specific induction signal can be compared by differential screening. A cDNA library is prepared from one of the two mRNA populations, and the two copies are immobilized on filters. The filters are then screened twice, once with highly labeled cDNA prepared from one of the two mRNA populations and once with a probe from the other mRNA population. By comparing the signals produced on the two filters probed with the different probes, it is possible to determine whether mRNA sequences are present in one population, but are absent or rare in the alternative mRNA population.

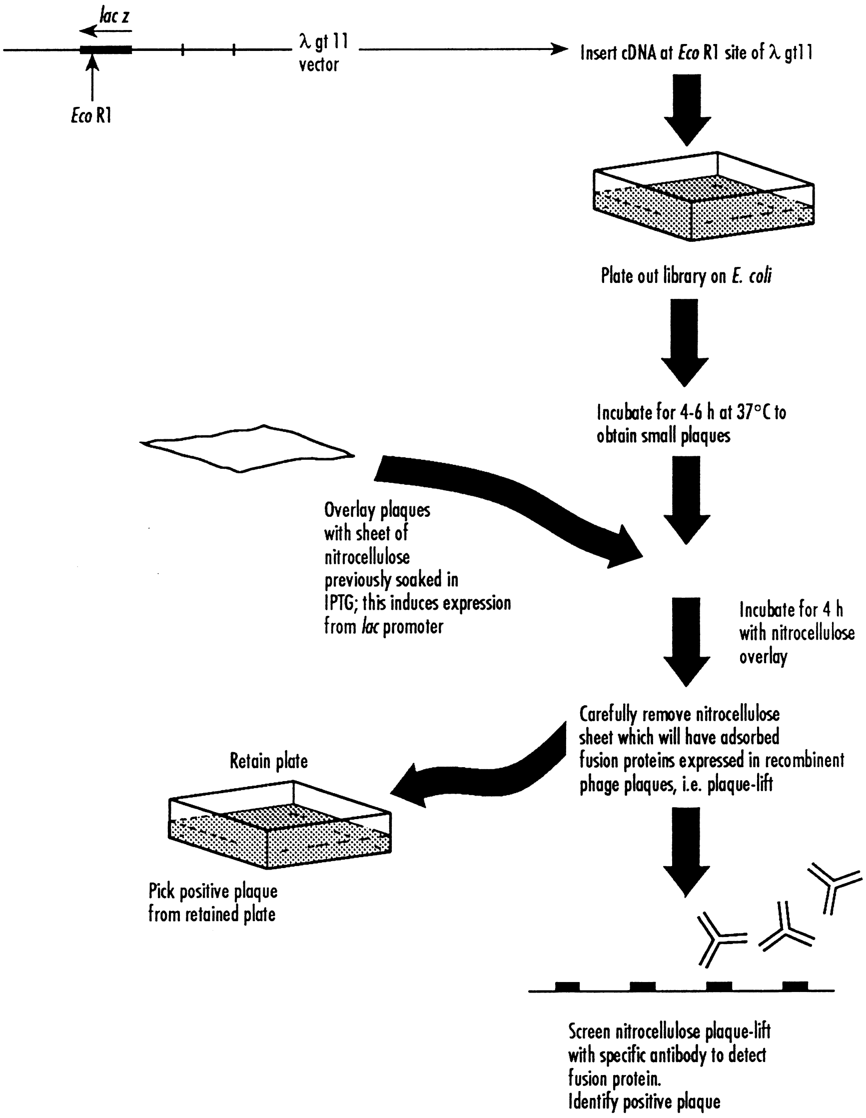


Figure 6.12. Identifying specific genes can also be done by immunochemical screening. A λ gt11 library in *E. coli* is screened in a manner similar to plaque hybridization except that the gene(s) of interest are identified by binding of a specific antibody to a fusion protein.

6.7. Labeling Probes by a Variety of Methods

Nucleic acid hybridization is used for many different purposes in molecular genetics. Nearly all phases of cloning and characterization or analysis of DNA involve hybridizing one strand of nucleic acid to another. Nucleic acid hybridization relies on the fact that two single-stranded nucleic acid molecules with complementary bases (DNA with DNA and DNA with RNA) are able to pair via hydrogen bonds. The strength of the hybridization is determined by the length of the homologous sequences, the experimental conditions, and the degree

of sequence homology. Chapter 5 described one application of nucleic acid hybridization, the Southern blot analysis, and one method for labeling probes, nick translation. Other nucleic acid hybridization techniques include colony or plaque hybridization and Northern blot analysis, in which the immobilized nucleic acid is RNA instead of DNA.

A number of labeling methods other than nick translation are available, as outlined below. The success of nucleic acid hybridization often relies on methods to introduce radioactive label into cloned segments of DNA or RNA. Each labeling technique has optimal sizes, efficiency, and different amounts of required nucleic acid template required. One measure of the efficiency of labeling is the specific activity of the label. Specific activity refers to the amount of radioactivity per microgram (μg) of probe DNA. The specific goals of the project will determine which labeling technique is employed. Detailed protocols are available in a variety of laboratory manuals and kits. The safe use of radioactivity for labeling requires that the user obtain specific training in handling procedures and disposal.

Nonradioactive probes using biotin and chemiluminescent labels also are available in kits and are safer for novices to use.

6.7.1. Synthesis of Uniformly Labeled DNA Probes by Random Primers

Short oligonucleotides can serve as primers for DNA synthesis by DNA polymerases on single-stranded templates. If the primers used are random in sequence, they will form hybrids at many different locations along the template strand so that the strand being synthesized will incorporate a radiolabeled dNTP randomly along its length. This ss DNA probe will have a very high specific activity. If reverse transcriptase is used for synthesis, the template can be RNA. If DNA is the template, then the Klenow fragment of DNA polymerase I is used.

6.7.2. Synthesis of Probes by Primer Extension

Primer extension is used to synthesize probes from denatured double-stranded DNA. It can be used to produce probes from denatured, closed circular DNA or from denatured, linear ds DNA. The purified DNA is mixed with random primers and denatured by boiling, and radiolabeled dNTPs and the Klenow fragment of DNA polymerase I are added to carry out synthesis of the probe. Random primers anneal to the denatured DNA, and the radioactive product is synthesized by primer extension. Probes prepared by random priming are usually 400 to 600 nt long.

6.7.3. End-Labeled Probes

A variety of methods are available to introduce labels at either the 3′ or 5′ ends of linear DNA. Usually only a single label is introduced at one end of the molecule, so the specific activities produced by such techniques are lower than those obtained by the uniform labeling methods described above. Both DNA and RNA can be end-labeled. The advantages to end labeling are that the location of the labeled group is known and very small fragments of DNA can be labeled, including restriction digest fragments.

6.7.4. Single-Stranded Probes

Single-stranded DNA, cDNA, or RNA probes have an advantage over ds probes because more probe is available to hybridize with the nucleic acid of interest. Single-stranded

probes should not anneal to themselves so that hybrids composed of reannealed probes cannot be made. RNA probes do not need to be denatured before being used because they are already single-stranded. As a result, RNA probes have higher specific activity (the ratio of radioactive to nonradioactive molecules of the same kind) than DNA probes. Double-stranded DNA probes must be denatured before using, which produces two strands. If only one of the DNA strands has been labeled, the unlabeled strand can dilute the reaction mixture.

Single-stranded probes are prepared from DNA templates by synthesizing radiolabeled DNA that is complementary to sequences cloned in a bacteriophage vector such as M13 or a phagemid. RNA probes can be produced by transcription of ds DNA in a vector with a powerful promoter derived from *E. coli* bacteriophages T7 and T3 by a bacteriophage DNA-dependent RNA polymerase. The labeled transcript produced is complementary to one of the two template strands. The probe can therefore be used as strand-specific probes in hybridization reactions. cDNA probes are used to isolate cDNA clones of genes that are expressed in specific cells or tissues.

6.7.5. Synthetic Probes

The knowledge of the sequence of a few amino acids in a protein will allow a specific gene to be isolated with a synthetic probe. Automated machines can synthesize short segments of ss DNA in which the sequences are defined precisely. Probes also can be produced that consist of alternative sequences, as determined by the degeneracy of the genetic code. Because there are 64 possible codons and only 20 amino acids, most amino acids are coded for by more than one codon. Thus, a probe that consists of a mixture of degenerate sequences can be used to screen libraries.

6.8. Baculovirus Vectors for Expressing Foreign Polypeptides in Insect Cells

E. coli, infected with plasmid or phage expression vectors, has been used to express foreign eukaryotic genes. However, it is sometimes difficult to obtain complex polypeptides derived from eukaryotes in a biologically active form because *E. coli* cannot make posttranslational changes to proteins such as **glycosylation** and **phosphorylation**. As a result, eukaryotic expression vectors have been developed for use in yeast and insect cells.

The most effective expression vectors used in insect cells were engineered from baculoviruses (Luckow and Summers 1988, Maiorella et al. 1988, Jaruis et al. 1990, O'Reilly et al. 1992). **Baculoviruses** are viruses with ds, circular DNA genomes contained within a rod-shaped protein coat. The Baculoviridae are divided into three subgroups: nuclear polyhedrosis viruses (NPV), granulosis viruses, and nonoccluded viruses.

Most NPVs primarily infect lepidopterans, where they produce nuclear inclusion bodies in which progeny virus particles are embedded. Polyhedrin is the protein component of the crystalline matrix that protects the viral particles when they are outside their insect host. Several NPVs have been used as biological pesticides in pest management programs. Perhaps the most extensively studied baculovirus is *Autographa californica* NPV. It has a relatively broad host range, and the life cycle of wild-type baculovirus begins when caterpillars eat the protein matrix (polyhedrin) which releases the virus particles. Virus replication occurs within host cells, but the protein matrix is not produced early in the infection. However, when the caterpillar is near death, the virus resumes polyhedrin production until approximately 20% of the insect cell proteins consists of polyhedrin.

The *Autographa californica* nuclear polyhedrosis virus (AcNPV) and the silkworm *Bombyx mori* nuclear polyhedrosis virus (BmNPV) have been exploited as vectors to carry exogenous DNA into insect cells in cell culture or into living silkworm larvae in order to produce foreign proteins (Maeda 1989). The productivity of baculovirus expression vectors is based on the extremely high efficiency of its polyhedrin gene promoter. The polyhedrin promoter enables very large amounts of the desired polypeptides to be produced. The level of expression of foreign gene products varies, depending on the specific foreign protein being produced.

Baculovirus expression vectors allow production of proteins that elicit an important biological response, but are often produced in tiny quantities in the normal host organism. A variety of mammalian proteins have been produced using *A. californica* NPV in cultured *Spodoptera frugiperda* cells. The first such protein, human interferon, was produced in 1983. Since then, many additional genes have been expressed, including growth factors, tumor antagonists, blood clotting or anticlotting factors, protein hormones, antibodies, and vaccines against hepatitis B, acquired immune deficiency syndrome (AIDS), and malaria.

In the next chapter, we learn how to sequence DNA and how to use these sequences.

General References

Ausubel, F. M., R. Brent, R. E. Kingston, D. M. Moore, J. G. Seldman, J. A. Smith, and K. Struhl, Eds. (1987). *Current Protocols in Molecular Biology*. Greene Publ. Assoc., New York. These two ring-bound volumes offer detailed research protocols. The publisher has a subscription update service so that new pages can be inserted to replace outdated material or expand on existing information.

Berger, S. L., and A. R. Kimmel, Eds. (1987). *Guide to Molecular Cloning Techniques*. Vol. 152, *Methods in Enzymology*. Academic Press, San Diego.

Brown, T. A., Ed. (1991). *Essential Molecular Biology. A Practical Approach*. Vols. I and II. IRL Press, Oxford.

Glover, D. M., Ed. (1985). *DNA Cloning Volume I—A Practical Approach*. IRL Press, Oxford.

Old, R. W., and S. B. Primrose (1989). *Principles of Gene Manipulation. An Introduction to Genetic Engineering*. 4th ed. Blackwell Scientific, Oxford, UK.

O'Reilly, D. R., L. K. Miller, and V. A. Luckow (1992). *Baculovirus Expression Vectors. A Laboratory Manual*. W. H. Freeman, New York.

Ptashne, M. (1986). *A Genetic Switch: Gene Control and Phage Lambda*. Blackwell Scientific, Oxford.

Sambrook, J., E. F. Fritsch, and T. Maniatis (1989). *Molecular Cloning: A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sambrook, J., and D. W. Russell (2001). *Molecular Cloning, A Laboratory Manual*. 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Volumes 1, 2, and 3 provide detailed protocols for a large number of techniques.

Williams, J. G., and R. K. Patient (1988). *Genetic Engineering*. IRL Press, Oxford.

References Cited

Jarvis, D. L., J. G. W. Fleming, G. R. Kovacs, M. D. Summers, and L. A. Guarino (1990). Use of early baculovirus promoters for continuous expression and efficient processing of foreign gene products in stably transformed lepidopteran cells. *Bio/Technol*. **8**: 950–955.

Kim, S. Y., S. K. Horrigan, J. L. Altenhofen, Z. H. Arbieva, R. Hoffman, and C. A. Westbrook (1998). Modifications of bacterial artificial chromosome clones using Cre recombinase: introduction of selectable markers for expression in eukaryotic cells. *Genome Res*. **8**: 404–412.

Kimmel, A. R., and S. L. Berger (1987). Preparation of cDNA and the generation of cDNA libraries: overview,. pp. 307–316. In: *Guide to Molecular Cloning Techniques*. Vol. 152, *Methods in Enzymology*. S. L. Berger and A. R. Kimmel, Eds. Academic Press, San Diego.

Luckow, V. A., and M. D. Summers (1988). Trends in the development of baculovirus expression vectors. *Bio/Technol*. **6**: 47–55.

Maeda, S. (1989). Expression of foreign genes in insects using baculovirus vectors. *Annu. Rev. Entomol*. **34**: 351–372.

Maiorella, B., D. Inlow, A. Shauger, and D. Harano (1988). Large-scale insect cell-culture for recombinant protein production. *Bio/Technol*. **6**: 1406–1410.

O'Reilly, D. R., L. K. Miller, and V. A. Luckow (1992). *Baculovirus Expression Vectors. A Laboratory Manual*. W. H. Freeman, New York.

Old, R. W., and S. B. Primrose (1989). *Principles of Gene Manipulation*. 4th ed., Blackwell Scientific, Oxford, UK.

Pouwels, P. H. (1991). Survey of cloning vectors for *Escherichia coli*. pp. 179–239. In: *Essential Molecular Biology. A Practical Approach*. Vol. I. T. A. Brown, Ed. IRL Press, Oxford, UK.

Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiri, and M. Simon (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**: 8794–8797.

Some Relevant Web Sites

Many Web sites are available with up-to-date information on protocols, equipment, and reagents. Most commercial suppliers maintain Web sites and electronic newsletters. Many protocols are found at university sites. For a starting point, try the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) or Biocompare, a buyer's guide for life sciences. This site is a source of catalogues, newsletters, and product reviews (http://www.biocompare.com/index.asp).

7

DNA Sequencing and the Evolution of the "-Omics"

7.1. Overview

Sequencing, resolving the order of the bases in DNA, is carried out on genomic DNA, cDNA, or mitochondrial DNA. Sequencing is a first step in evaluating regulatory sequences as well as coding and noncoding regions. DNA sequences are used to reconstruct phylogenies, identify taxonomic groups, evaluate the evolution of genomes, and conduct research on population ecology and genetics.

There are two original methods for sequencing DNA. For both, sequencing involves four procedures: 1) cloning/preparing template DNA, 2) performing the sequencing reactions, 3) gel electrophoresis of the samples, and 4) compiling and interpreting the data.

The most commonly used sequencing method, the Sanger or dideoxy chain terminating method, was developed in 1975. This method involves synthesis of DNA *in vitro* on a single-stranded template using a primer, a mixture of labeled deoxynucleotide triphosphates (dNTPs) and dideoxynucleotide triphosphates (ddNTPs). The reaction terminates at the position at which a ddNTP, instead of an dNTP, incorporates into the growing DNA chain. Four different reactions are carried out, one for each base. The DNA fragments from the four reactions are separated by acrylamide gel electrophoresis, and the base sequence is identified by autoradiography of the four banding patterns.

In the "chemical" sequencing method developed by Maxam and Gilbert in 1977, single-stranded DNA, derived from double-stranded DNA and labeled at the $5'$ end, is subjected to chemical cleavage protocols that selectively make breaks on one side of a specific base. Fragments from the reactions are then separated according to size by acrylamide gel electrophoresis, and the sequences are identified by autoradiography.

DNA sequencing now can be highly automated, using bases labeled with fluorescent dyes. Sequences can be read by scanners directly, and sequence analysis is automated, which has reduced costs and increased speed. Automated sequencing methods were used for the very large-scale sequencing required to sequence the entire *Drosophila melanogaster* genome. The information obtained from the *Drosophila* Genome Project has begun to revolutionize both fundamental and applied aspects of insect genetics.

As a result of the genome projects, several new scientific disciplines, called "-omics," are being developed. Genomics is providing insight into development, speciation, protein interactions, and evolution. Proteomics involves understanding the structure and function of the proteins encoded by the genes. Transcriptomics resolves which genes are transcribed during development in specific tissues. Phenomics attempts to understand how the genes and proteins interact in physiological cascades to determine the phenotype and development of organisms. Because genomics and proteomics produce huge amounts of data, biologists need to use computers and other information management tools, which has generated a new discipline called bioinformatics.

7.2. Introduction

DNA sequencing is an important component of many molecular genetics projects. Sequencing often is a necessary component of a project, while in other cases it is the desired end point and the sequences are used in taxonomic, ecological, or evolutionary studies.

Advances in technology have made it feasible to sequence entire genomes, which is revolutionizing both basic and applied knowledge of gene structure, gene function, and evolution. In identifying the sequences of promoters, protein coding sequences, and noncoding regions of DNA in genomic or mitochondrial DNA, it is possible to deduce relationships between organisms and reconstruct their evolutionary history.

The development of extensive computerized databases of DNA and protein sequences allows hypotheses to be constructed and tested regarding the structure and function of proteins and their secondary structures. All of these opportunities became possible only after DNA sequencing methods were developed in the late 1970s.

Two basic DNA sequencing methods were developed at about the same time: the chemical or Maxam–Gilbert method (1977, 1980) and the chain-terminating method of Sanger et al. (1977). Both utilize the same basic approaches: 1) cloning or preparing the DNA templates, 2) performing the sequencing reactions on the DNA templates, 3) gel electrophoresis of the samples, and 4) compiling and interpreting the data.

DNA to be sequenced can be genomic DNA, mitochondrial DNA, or cDNA. Because cDNA lacks introns and regulatory elements, sequencing of cDNA provides less information. Sequencing only cDNA probably would miss some genes that are expressed at very low levels or in a tissue- or time-dependent manner.

Effective computer tools are needed to discover the sequences that actually code for a gene, because up to 90% of genomic sequences are noncoding DNA. Some noncoding DNA sequences are associated with centromeres or telomeres, and others have no known function. Different computer programs have been developed to search DNA sequence data and identify possible regulatory sequences, potential start or stop codons, open reading frames (ORFs), and sequences that may indicate the location of intervening sequences or introns. Unfortunately, no computer program at present is 100% accurate in identifying genes (Bork 2000).

The most reliable way to find genes currently is to identify them because they are similar to known proteins from the same or other organisms or similar to cDNAs from the same or a closely related organism (Stormo 2000). However, many genes have no significant similarity with other known sequences.

The length of DNA that can be sequenced by a single reaction varies from 200 to ~1000 bases, depending upon the method employed (Sambrook and Russell 2001). Vectors can contain DNA inserts ranging in size from 100 to 1,000,000 bp. For example, yeast artificial chromosomes (YACs) can contain inserts up to 1 million bp long, and cosmids can contain inserts 30,000 to 45,000 bp long. Thus, cloned DNA typically is converted into smaller segments or **subclones**, which are then inserted into vectors that are specialized for sequencing, such as M13 or plasmid sequencing vectors.

7.3. The Dideoxy or Chain-Terminating Method

Briefly, the **dideoxy** or **chain-terminating sequencing** method developed by Sanger involves *de novo* synthesis of a series of labeled DNA fragments from a single-stranded (ss) DNA template. Two methods are employed to produce a ss DNA template: 1) denaturing double-stranded (ds) DNA by heating it, and 2) cloning it into a vector that produces ss DNA. The ss DNA segment to be sequenced serves as the template for the synthesis, by complementary base pairing, of a new labeled strand of DNA. Labeling initially was achieved by labeling with ^{32}P or ^{35}S.

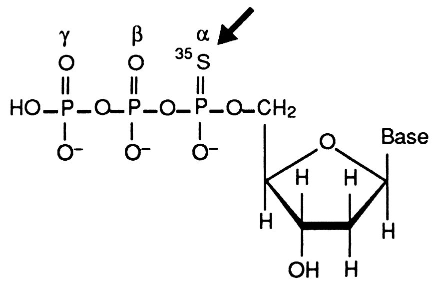


Figure 7.1. Structure of α-^{35}S-deoxynucleoside triphosphate. Labeling with ^{35}S results in sharp bands, which increases the resolution of the sequencing gels.

Because DNA synthesis is employed in the chain-terminating method, the sequencing reaction requires a DNA polymerase, labeled deoxyribonucleotides (dNTPs), a primer, and dideoxyribonucleotides (ddNTPs) (Figure 7.1). Several different DNA polymerases could be used (Klenow fragment, Sequenase 2, or thermophilic DNA polymerases such as *Taq*), with different protocols (Sambrook and Russell 2001). The dNTPs can be labeled either with ^{32}P or ^{35}S, but ^{35}S labeling produces sharper bands and improves the resolution of the autoradiogram. The structure of a dNTP that has been labeled with ^{35}S is shown in Figure 7.2.

DNA sequencing kits that contain the necessary enzymes and components can be purchased. Each has specific protocols provided and are ideal for first-time sequencers, although they are too expensive for large-scale projects (Sambrook and Russell 2001).

The dideoxy or chain-terminating reaction is begun by adding a short oligonucleotide primer that is complementary to a region of DNA adjacent to the DNA segment to be sequenced (Table 7.1, Figure 7.2). The primer is normally 15 to 30 nucleotides long and is annealed to the template in a preincubation step. Four separate reactions are set up to determine the position, respectively, of the A, T, G, and C bases in the template DNA. Each reaction requires a mixture of DNA polymerase, primers, dNTPs, ddNTPs, and the template DNA. The ddNTPs are derivatives of dNTPs that do not contain a hydroxyl group at the 3′ position of the deoxyribose ring (Figure 7.3).

When ddNTPs are incorporated into the growing DNA chains instead of dNTPs, that particular DNA molecule is terminated at that point. All four dNTPs are present in each tube with each ddNTP, but the ratio is adjusted so that ddNTPs are less frequent than dNTPs. This makes the incorporation of a ddNTP a random event. The newly synthesized DNA molecules in a specific reaction tube therefore are a *mixture* of DNA fragments of different lengths, each with a fixed *starting point* (determined by the primer) but with variable ending points. Thus, for example, in the reaction in which the chain is terminated when thymines (T) are incorporated, a ddNTP does not always get incorporated into the first site where a T occurs. Nor does a ddNTP necessarily get incorporated into the chain where the second T occurs. However, over the length of the DNA being sequenced, each site where a T is incorporated will have ddNTPs incorporated, so that a series of DNA molecules of different length is produced (Figure 7.4). Thus, *populations* of synthesized molecules are produced in which the chain is terminated at each site where Ts occur. These DNA molecule populations can be visualized by gel electrophoresis.

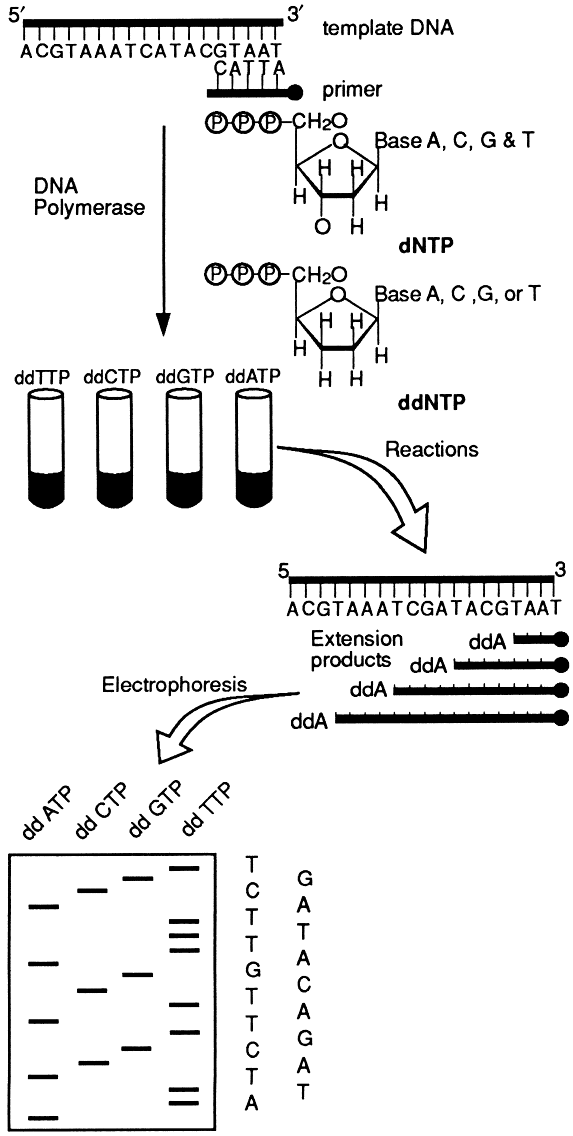


Figure 7.2. A diagram of the steps involved in the Sanger dideoxy chain-terminating method of DNA sequencing (modified from Hunkapiller et al. 1991). Four reactions are carried out.

The base sequence is visualized by running the radiolabeled DNA fragments from the four reactions on an acrylamide gel in four adjacent lanes. Each reaction tube will produce a series of bands, with each band representing a population of molecules at which the DNA molecule was terminated by incorporating a ddNTP. The banding pattern in the four lanes is visualized on an X-ray film (Figure 7.4). The two strands should be sequenced independently to reduce errors generated by artifacts in the sequence reactions or inadequate resolution of regions of the sequence on the gel.

Table 7.1. An Example of a Dideoxy (Sanger) Sequencing Protocol Using Modified Bacteriophage T7 DNA Polymerase (Sequenase)

Sequencing Reagents

Annealing Buffer (5× concentrate)

200 mM Tris·HCl pH 7.5
100 mM $MgCl_2$
250 mM NaCl

Dithiothreitol (DDT) 0.1 M

Labeling nucleotide mixture (for use with radiolabeled dATP)

1.5 mM dGTP, 1.5 mM dCTP, and 1.5 mM dTTP

Termination nucleotide mixtures (one for each dideoxynucleotide)

Each mixture contains 80 mM dGTP, 80 mM dATP, 80 mM dTTP, 80 Mm dCTP, and 50 mM NaCl. In addition, the "G" mixture contains 8 mM dideoxy-dGTP; the "A" mix, 8 mM ddATP; the "T," 8 mM ddTTP; and the "C," 8 mM ddCTP.

Stop solution: 95% Formamide, 20 mM EDTA, 0.05% bromphenol blue, and 0.05% xylene cyanol FF

Procedures

Annealing Template and Primer

1. For each template, a single annealing (and subsequent labeling) reaction is used. Combine the following:

Primer	0.5 pmol
DNA	0.5–1.0 pmol (1 mg of M13)
Annealing buffer	2 ml
Water to bring the total volume to 10 ml	

Warm the capped tube to 65°C for 2 min, allow the mixture to cool slowly to room temperature over a period of about 30 min.

Labeling Reaction

2. To the annealed template-primer add the following

DTT (0.1 M)	1 ml
Labeling nucleotide mix	2 ml
α-^{35}S or α-^{32}P dATP	5 mCI (typically 0.5 ml)
Sequenase™	3 units

Total volume should be approximately 15 ml; mix thoroughly and incubate for 2–5 min at room temperature.

Termination Reactions

3. Label 4 tubes "G," "A," "T," and "C." Fill each with 2.5 ml of the appropriate dideoxy termination mixture.

When the labeling reaction is complete, transfer 3.5 ml of it to the tube prewarmed to 37°C, labeled "G". Do the same for each of the other three tubes (A, T, and C).

After 2–5 min at 37°C, add 4 ml of Stop Solution to each termination reaction, mix, and store on ice. To load the gel, heat the samples to 75–80°C for 2 min or more and load 2–3 ml in each lane.

4. Polyacrylamide gel electrophoresis

5. Analysis of sequences

Modified from Tabor and Richardson (1987).

7.4. Variations on Dideoxy Sequencing Methods

There are a number of different protocols for sequencing DNA by the dideoxy chain termination method (Ambrose and Pless 1987, Barnes 1987, Mierendorf and Pfeffer 1987,

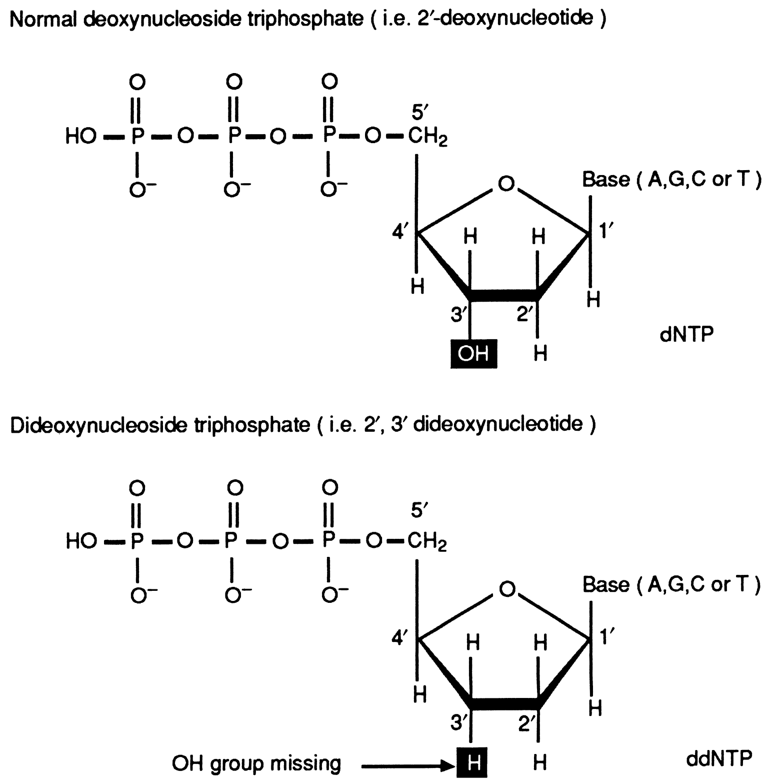


Figure 7.3. Dideoxynucleoside triphosphates (ddNTPs) (bottom) act as chain terminators because they lack a 3′-OH group found on normal deoxynucleotides (dNTPs) (top).

Howe and Ward 1989, Sambrook et al. 1989, Tabor and Richardson 1987). Variables include such factors as whether the DNA to be sequenced is double- or single-stranded and whether the DNA is sequenced directly from ds plasmid DNA or after subcloning into a ss phage such as M13.

7.5. DNA Sequences Can Be Analyzed on Polyacrylamide Gels

Both DNA sequencing methods generate sets of DNA molecules that share a common end that is determined by the primers, but vary in their length at the other end. Both methods also originally employed radioactive labeling to visualize the results. Once the DNA is electrophoresed, the gel is dried onto paper and exposed to X-ray film. The autoradiogram produced displays a ladder of bands representing the migration position of the different radiolabeled DNA segments. For example, the sequence of 20 nucleotides can be read to the right of the four lanes in Figure 7.4.

Both sequencing methods originally used very thin polyacrylamide gels to discriminate between nucleic acid molecules that may differ in length by only one nucleotide. A sequencing gel is a high-resolution gel containing 6 to 20% polyacrylamide and 7 M urea. The DNA to be analyzed is denatured before electrophoresis by heating it to 80°C in a

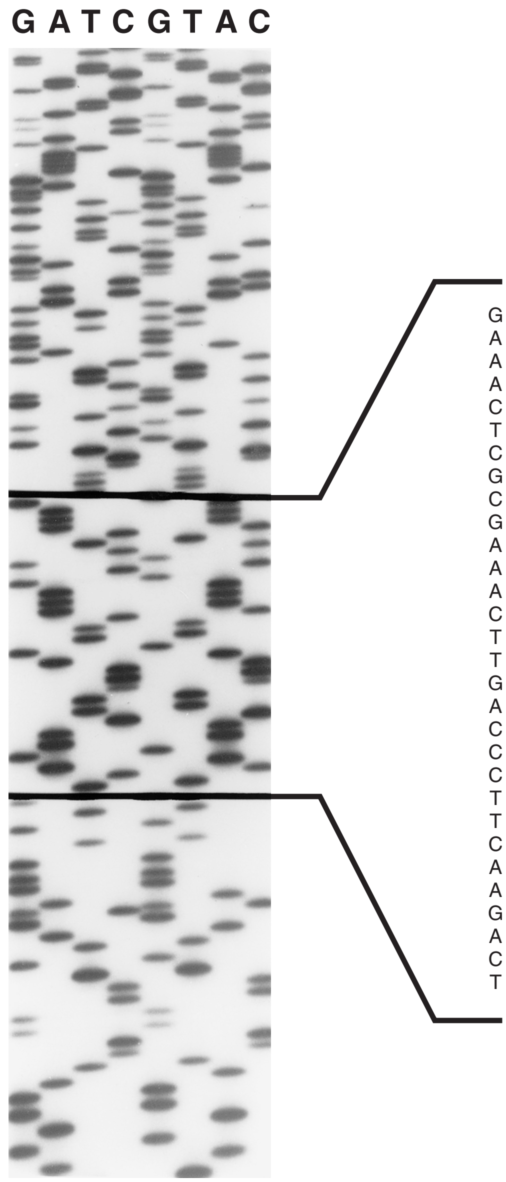


Figure 7.4. Autoradiograph of a sequencing gel obtained by the dideoxy chain-terminating method. Four different reactions are carried out, in which DNA synthesis of a complementary DNA chain is terminated by incorporating a radiolabeled dideoxy base (ddATP, ddCTP, ddGTP, or ddTTP) rather than a deoxy base (dATP, dCTP, dGTP, dTTP). The resulting fragments of synthesized DNA are visualized by acrylamide gel electrophoresis and autoradiography. The sequence is read by determining which lane contains each succeeding DNA segment, as determined by a band. Thus, reading *from the bottom*, the shortest fragment has a band in the T lane (just above the line). The next band is a C, then the next is an A, and so on. Bands at the very top and bottom of the gel are not read. This sequence is a portion of the *mariner* transposable element cloned from the predatory mite *Metaseiulus occidentalis* (Jeyaprakash and Hoy 1995).

buffer containing formamide, which lowers the melting temperature of ds DNA molecules. The urea minimizes DNA secondary structure, which could affect mobility of the DNA through the gel. The gel is run using sufficient power so that it is heated to about 50°C, which also minimizes DNA secondary structure. The rate of migration of an individual DNA molecule is determined by its length. The shorter fragments migrate more quickly than the longer, thus ending up nearer the bottom of the gel. Protocols for gel electrophoresis of DNA are widely available (Davies 1982, Sealey and Southern 1982, Howe and Ward 1989).

7.6. Sequencing Reactions Require a Primer

If the template DNA is a subclone that was inserted into the multiple cloning site of a vector, then the primer almost always is an oligonucleotide complementary to sequences flanking the multiple cloning site. This allows any fragment cloned into the multiple cloning sites to be sequenced using the same primer. Most vectors have the *lacZ* sequences flanking the multiple cloning sites. Thus, an oligonucleotide primer directed to this sequence, which is 16 to 17 nucleotides long, is commonly used and is called the **M13 universal primer**. This primer is designed to be complementary to the strand of DNA packaged into M13 or into any plasmid vector containing an M13 origin of replication. Methods also have been developed for direct sequencing from denatured plasmid DNA (Mierendorf and Pfeffer 1987), which eliminates the need to isolate or subclone DNA fragments.

7.7. The Maxam and Gilbert Sequencing Method

The **Maxam and Gilbert DNA sequencing** method also is called the "chemical-cleavage" method (Maxam and Gilbert 1977, 1980). It uses chemical reagents to generate base-specific cleavages of the DNA to be sequenced (Figure 7.5, Table 7.2). It is less used today, in part because the chemicals used are toxic and the methods are labor intensive. The primary advantage of this method is that DNA sequences are obtained from the original DNA molecule and not from a synthesized copy. Thus, one can analyze DNA modifications such as methylation and study DNA secondary structure and the interaction of proteins with DNA.

To start, one needs pure DNA that has been cut by restriction endonucleases to generate DNA of specific length and with known sequences at one end. Each DNA fragment then can be radioactively labeled at one end with a ^{32}P-phosphate group in sufficient quantity that at least four different chemical reactions can be carried out.

Next, specific bases in the DNA fragment are altered in at least four separate chemical reactions (Table 7.2). For example, guanine (G) is methylated by dimethyl sulfate. Each reaction is carried out in a manner that limits the reaction so that, for example, only one G is modified per several hundred G nucleotides. The altered G is then removed, in a subsequent step, by cleavage at the modification points with piperidine. The result is a set of end-labeled fragments of different lengths that will show up as a ladder of bands on the gel because the reaction was limited and not all the Gs were altered in the reaction. Four different reaction samples (one for $A > C$, G, $T + C$, and C) are then run side by side on a sequencing gel and the results can be visualized by autoradiography (Figure 7.5).

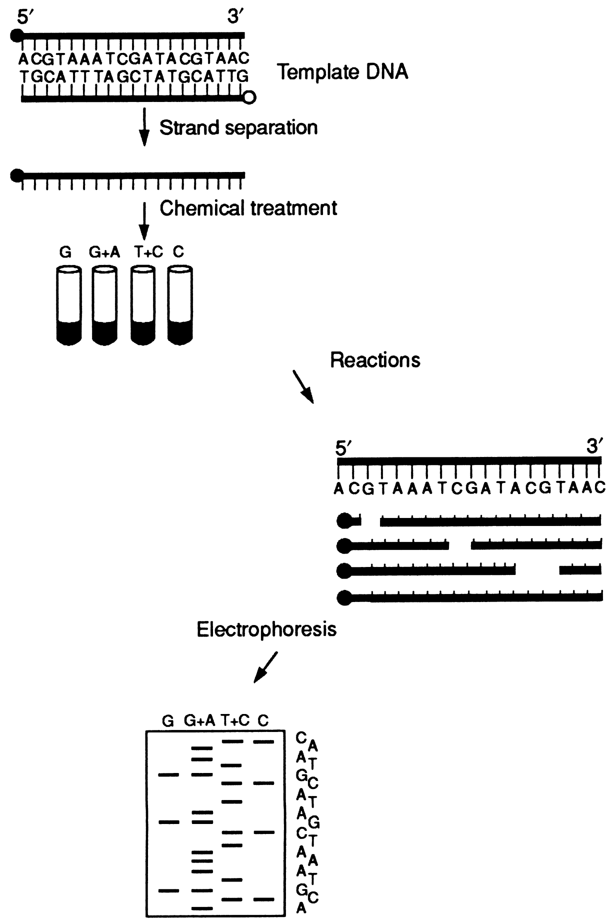


Figure 7.5. A diagram of the steps involved in the Maxam and Gilbert chemical cleavage method of DNA sequencing (modified from Hunkapiller et al. 1991).

Table 7.2. Reagents for Maxam and Gilbert DNA Sequencing

Base specificity	Base-modifying reagent	Base-modifying reaction
G	Dimethyl sulfate	Methylation of N7 makes C8–C9 bond susceptible to cleavage by base
G + A	Formic acid	Weakens bond of A + G
T + C	Hydrazine	Opens pyrimidine rings which make them susceptible to removal
A	Hydrazine + NaCl	Opens cytosine rings

7.8. Shotgun Strategies for Genomes

In "shotgun" sequencing strategies, the DNA is digested with a restriction endonuclease, and subfragments are cloned and sequenced (Figure 7.6B). The nucleotide sequence of

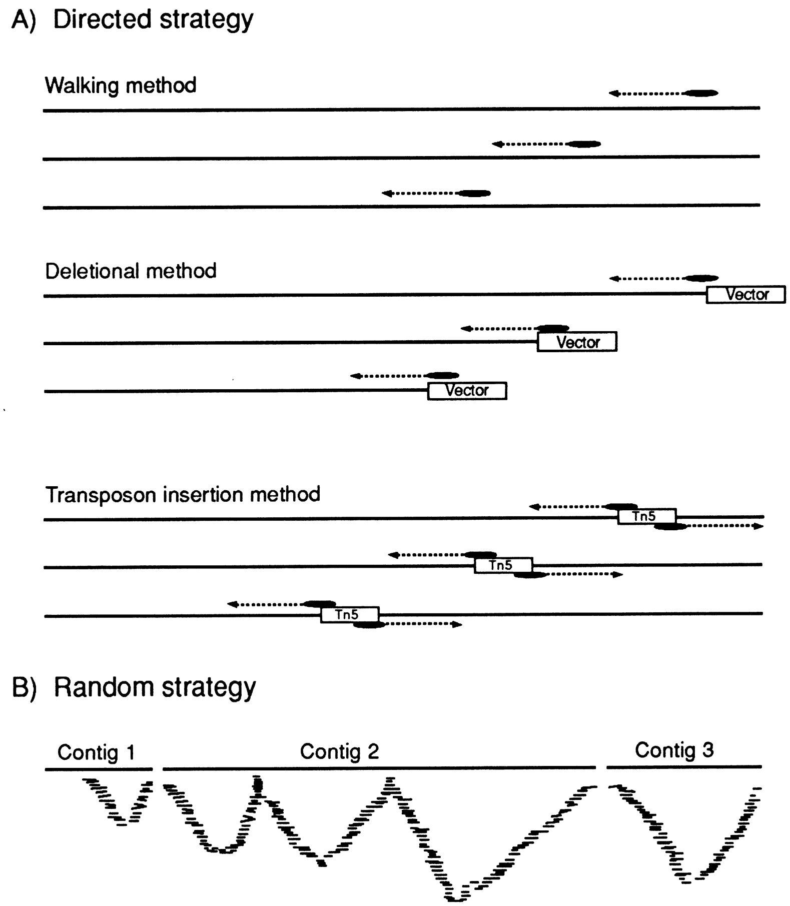


Figure 7.6. Strategies for DNA sequencing of long DNA segments involve either directed (A) or random (B) methods. Directed methods include walking, deletional sequencing, or transposon insertion. Analyzing sequences from many contigs requires large-scale computer alignments.

the various inserts is obtained, and a computer is used to determine how the fragments (= **contigs**) overlap and to establish the entire sequence of the original DNA fragment used to generate the subclones.

Disadvantages to the shotgun method are that it may underrepresent some fragments, sequencing must be redundant to ensure that the entire sequence has been included in one or more subclones, and there is no way to identify specific fragments.

7.9. Sequencing DNA by the Polymerase Chain Reaction (PCR)

The **polymerase chain reaction** (PCR) is a procedure by which a specific segment of DNA can be amplified by one millionfold, or more, using a DNA polymerase (see Chapter 8). DNA fragments can be amplified by the PCR directly from genomic or cloned DNA,

which eliminates the need to prepare large amounts of DNA from tissues for sequencing and subcloning steps.

Conventional PCR requires primers of known sequence that flank the region to be amplified. Several techniques have been developed for sequencing ds DNA produced by the PCR (Chapter 8). Cycle sequencing is advantageous because it requires only very small amounts of template DNA, and it can work with ds as well as ss templates, which eliminates the need to subclone the DNA into M13 or phagemid vectors.

7.10. Automated DNA Sequencers

The invention of automated fluorescent DNA sequencers made large-scale genome projects feasible. A variety of instruments can automate nearly every step of the large-scale sequencing process. Integrated machines can isolate DNA, clone or amplify DNA, prepare enzymatic sequencing reactions, purify DNA, and separate and detect DNA fragments containing fluorescent labels to obtain the DNA sequence (Meldrum 2000a).

Automated sequencers use horizontal or vertical slab gels, and some recent commercial systems can use capillary sequencers (Meldrum 2000b). A commonly used sequencer, the ABI PRISM 377 DNA Sequencer, uses multicolor fluorescence labeling and a four-dye, one-lane detection system. Two hundred bases per sample per hour can be analyzed, and 18, 36, 64, or 96 samples can be analyzed simultaneously on vertical gels. The gel plates come in four different lengths to optimize run times and sample resolution.

Many large-scale sequencing facilities use a random shotgun phase combined with a directed finishing phase to complete analysis of the difficult regions of the genome. Others use a whole-genome shotgun approach in which random fragments of total genomic DNA are subcloned and high-throughput sequencing is used to generate sequences that provide at least a 10-fold coverage of the genome. These sequences are ordered and put into a linear sequence with the aid of very high speed computers (Meldrum 2000a).

As a result of the industrialization of DNA sequencing within the past few years, the cost has decreased to approximately $0.20 to 0.30 per base when the accuracy is held to less than one error in 10,000 bases. As a result of these reduced costs, there has been an exponential growth in the number of sequencing projects being carried out.

During the year 2000, approximately one complete bacterial genome was obtained each month. The *Drosophila* Genome Project was completed in the fall of 2000 (Adams et al. 2000), and on June 26, 2000, a working draft of the human genome was completed. The year 2000 truly can be called "The Year of the Genome."

7.11. Analyzing DNA Sequence Data

DNA sequencing projects generate substantial amounts of data and require computer assistance for their analysis (Howe and Ward 1989, Doolittle 1990, Weir 1990, Gribskov and Devereux 1991, Reese et al. 2000, Stein 2001). Software packages are available for all common laboratory computer systems and, depending on the size of the computer, can analyze the sequences in greater or lesser detail. Sequence data can be put into the computer in several ways. Gels can be read visually by the scientist, digitized and entered, or scanned by automated laser scanners which can enter the data directly. DNA sequences obtained from

automated sequencing machines are usually provided online or on computer disk. Software has been developed that can interpret ambiguities in the sequence data.

The automated methods provide speed and help to minimize clerical errors that can occur if data are entered manually. Computer programs can compare readings from several sequencing runs, search for and identify overlaps, compare results from sequencing the complementary strands of the DNA, and identify possible errors. Once the sequences have been entered into the computer and possible clerical or experimental errors resolved, the next step is to analyze the data.

In a random or shotgun sequencing project, the DNA is broken into fragments, which are cloned and sequenced. The relationships between the cloned fragments are determined by comparing their sequences. DNA segments that are related to one another by a partial overlap are called **contigs**. If a sequence overlaps with another, then the two contigs can be joined. The process of comparing sequences and aligning them is continued until it is possible to produce a continuous DNA sequence for the DNA of interest.

A variety of questions can be asked after the sequences are obtained. The sequence can be searched for all known restriction endonuclease target sites, and the computer can generate a comprehensive and precise restriction map. The sequences can be searched for interesting structures such as **tandem repeats** and **inverted repeats**, which would indicate the insertion of transposable elements. The sequences can predict which proteins are coded for based on the sequences in the possible open reading frames on each of the two strands. An **open reading frame (ORF)** is a segment of DNA that does not include a termination codon and may contain a polypeptide coding region. Both strands must be interpreted because it is not known in advance which is the coding and which is the noncoding strand.

The DNA sequence itself, or the deduced polypeptide sequence, may be compared with sequences in data banks. Often, because of the degeneracy of the DNA code, similarities are found when two polypeptide sequences are compared; these sequence similarities might not have been apparent if the comparison had been carried out only at the DNA sequence level.

DNA sequence similarities may be present because of **convergent evolution** or through homology. Convergent evolution implies that the two sequences did not have a common ancestral sequence, but that selection for a particular function in two different lineages has converged on a particular structure or related structures.

The term **homology** has become controversial because it has multiple definitions. Traditionally, similar structures in different organisms have been called homologous if the organisms have descended from a common ancestor. However, some molecular biologists have used the term "percent homology" when they mean there is *similarity* in DNA sequence, which may not be due to descent from a common ancestor.

A search of the DNA or protein sequence banks for **similarities** with any newly discovered sequence may turn up amazing degrees of similarity. For example, in the fruit fly *D. melanogaster*, **homeotic** genes, genes that direct cells in different segments to develop in particular patterns, have been cloned. The **homeobox**, a segment of approximately 180 bp, is characteristic of the homeotic class, and probes using the homeobox sequence have been used to isolate previously unknown homeotic genes in other insects. Even more interesting, sequences homologous to the homeobox have been isolated from mice and humans, indicating that similar genetic mechanisms may control some aspects of development in higher organisms. The high degree of conservation between the homeobox sequences of *Drosophila*, frogs, mice, and man indicates that these sequences have been

conserved for more than 500 million years, which is when invertebrates and vertebrates are thought to have diverged.

7.12. DNA Sequence Data Banks

DNA sequence data banks are expanding rapidly and are important resources for the research community. There are three major DNA sequence databases: the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory Nucleotide Sequence Data Library (EMBL), and the GenBank Genetic Sequence Data Bank (GenBank). Subsets of the databases have been organized. For example, there is a database of metazoan mitochondrial DNA sequences (Lanave et al. 2000), a eukaryotic promoter database (Perier et al. 2000), a database of restriction enzymes and methylases (Roberts and Macelis 2000), a database for intron sequence and evolution (Schisler and Palmer 2000), a database for homeodomains (Banerjee-Basu et al. 2000), and many others (Wheeler et al. 2000).

The data banks can be searched over the Internet, but some caution is required because, although efforts are made to ensure that the data entered are accurate, errors apparently are common in both the data banks and scientific journals. Errors, particularly in noncoding regions, may arise from sequencing or clerical errors. Submission of a sequence to a database in a machine-readable form is becoming a prerequisite for publishing in many journals and can be accomplished with electronic computer mail systems (Cinkosky et al. 1991). Methods of DNA sequence analysis using computer programs such as BLAST and PAUP will be described in Chapter 12 (Molecular Systematics and Evolution of Arthropods) and are described by Mount (2001).

7.13. A Brief History of the *Drosophila* Genome Project

The Human Genome Project was the reason the *Drosophila* Genome Project was begun. The Human Genome Project was one of the largest initiatives in the history of biology and at first was one of the most controversial. To prepare for this enormous undertaking, the genomes of the yeast *Saccharomyces cerevisiae*, the nematode *Caenorhabditis elegans*, the fruit fly *D. melanogaster*, and the laboratory mouse *Mus musculus* were targeted for sequencing in order to develop and improve methods prior to undertaking the larger human genome. Even these proposed subprojects generated considerable controversy.

Controversy arose as to whether it was appropriate to spend the time and resources to sequence the *Drosophila* genome. Knowledge of the structure and function of the *D. melanogaster* genome already was far greater than that for any other multicellular organism (Kafatos et al. 1991), and some believed a *Drosophila* Genome Project was unnecessary. The *D. melanogaster* genome is approximately 180 Mb of DNA, a third of which is heterochromatin. The 120 Mb of euchromatin is on the two large autosomes and the X; the fourth chromosome is mostly heterochromatin, with only about 1 Mb of euchromatin. By 1991, approximately 3800 different genes, about one-fourth of the total, already had been mapped by recombination studies. Many had been associated cytogenetically with one of the 5000 bands of the polytene salivary gland chromosomes. Approximately 3000 "transcription units" had been placed on the cytogenetic map by localizing the DNA on specific polytene chromosomes by *in situ* hybridization. Nearly 10% of the total genes,

1300 genes, in *D. melanogaster* already had been cloned and sequenced by individual laboratories (Rubin and Lewis 2000). However, eventually it was concluded that approximately one-third of the genes in *Drosophila* do not have obvious phenotypes when mutated, making the sequencing project a useful gene discovery method.

7.13.1. The Original *Drosophila* Genome Project

The original publicly supported *Drosophila* Genome Project had the following aims:

1. Develop a high-resolution **physical map**, which would serve as a basis for DNA sequencing and detailed functional studies. A physical map is a series of overlapping clones for which information is available on the sequences at their ends and on their physical location on the chromosomes. The physical map would be integrated in a database with cross-references to the genetic information already available for *D. melanogaster*.
2. Conduct feasibility studies for large-scale DNA sequencing projects, especially for regions containing DNA of great biological interest. Large-scale studies were defined as those that attempted to determine 3 megabases of contiguous DNA sequence within 3 years.
3. Develop new bioinformatic techniques to identify coding sequences in genomic DNA and to obtain high-quality cDNA libraries that were representative of the complete coding information of the genomic DNA (Merrian et al. 1991).

7.13.2. The Actual *Drosophila* Genome Project

The *Drosophila* Genome Project actually was completed much more quickly and by a different strategy than originally planned (Adams et al. 2000, Pennisi 2000a). *Drosophila melanogaster* became only the second multicellular organism (after the worm *C. elegans*) to have its entire genome sequenced.

The initial *Drosophila* sequencing effort was initiated in 1990 and was only partially completed when Venter et al. (1996) proposed using a "shotgun strategy." This was a novel approach to sequencing such a large genome: it involved breaking the entire genome into small pieces, sequencing them rapidly with a huge array of very fast and expensive new sequencing machines, then using some of the world's most powerful supercomputers to assemble the sequenced fragments into the correct order. To prove that this strategy would work, a collaboration was undertaken by a company founded by Craig Venter (Celera), the Berkeley *Drosophila* Genome Project, and its European counterpart to guide the work and interpret the data.

Shotgun cloning had never been attempted previously with such a complex genome. The complexity is due to the presence of repeated sequences hundreds to thousands of base pairs long that are scattered throughout the genome and cause problems in assembly of the sequence data. The solution was to obtain sequences from both ends of fragments that were approximately 2, 10, and 150 kb in length. These oriented bits of sequence were assembled into increasingly dense and interlinked scaffolds that generated long continuous stretches of DNA sequence with few gaps (Adams et al. 2000). The success of the shotgun approach with *Drosophila* encouraged a similar approach with the human genome (Venter et al. 2001).

The approach to sequencing the *Drosophila* genome began in May 1999 at Celera, and by late fall of 1999 the sequencing was completed and the computers had assembled the sequences! The sequences were published in *Science* in March of 2000 (Adams et al. 2000).

The publication of the *Drosophila* genome represents a major milestone for insect molecular genetics (Hawley and Walker 2000). The entire *Drosophila* sequence is available in GenBank and at FlyBase on the World Wide Web (FlyBase Consortium 1997). FlyBase is a database of genetic and molecular data and includes genes, alleles, phenotypes, aberrations, transposons, clones, stock lists, the locations of *Drosophila* workers, and bibliographic references (Misra et al. 2000).

By the year 2000, the genomes of *E. coli*, *Saccharomyces*, *C. elegans*, *D. melanogaster*, and *Homo sapiens* had been completed ahead of schedule and less expensively than expected. On February 15, 2001, the sequences of the human genome were published (International Human Genome Sequencing Consortium 2001, Venter et al. 2001).

7.13.3. Genome Analysis

Obtaining the DNA sequence is only a first step (Stein 2001). Analyses of the sequence data must be conducted. An early analysis involved "annotating" the *Drosophila* genome, which means that as many genes as possible were identified in the mass of data and the function of the proteins/gene products were predicted (Adams et al. 2000, Reese et al. 2000, Pennisi 2000b).

The accuracy of the annotation was assessed by several methods in the Genome Annotation Assessment Project (GASP). GASP focused on analysis of a well-known region of the *Drosophila* genome. The 12 groups carrying out the analysis did best in identifying the coding regions, with a success rate averaging over 95%. The correct intron/exon structures were predicted for over 40% of the genes. Almost half the genes in the region were recognized and assigned functions by homology with known genes. However, promoter analyses were highly inaccurate, and fewer than one-third of the promoters in the region were found by the GASP group (Reese et al. 2000). Subsequent annotations and evaluations are required to refine and improve on these initial annotations.

7.13.4. Surprises in the *Drosophila* Genome

Several unexpected results were found. First, early analyses of the *Drosophila* genome suggest there are approximately 13,600 genes, which is slightly fewer than the number found in the nematode *C. elegans* (Adams et al. 2000). *D. melanogaster* was expected to have about 30,000 genes. Ultimately, *Drosophila* may be found to have more than 13,600 genes because it has a relatively large number of overlapping genes (Ashburner 2000).

Immediately after obtaining the *D. melanogaster* genome sequences, a comparison was made to the genomes of *C. elegans* and *S. cerevisiae* in the context of cellular, developmental and evolutionary processes (Rubin et al. 2000). The comparisons indicated there are many genes left to be studied in *Drosophila*. For example, there are approximately 3090 genes encoding translation factors or proteins with well-characterized RNA binding motifs (Lasko 2000); there are approximately 251 protein kinases and 86 phosphatases in the genome, among which there are almost 170 new kinases and phosphatases for which no role is known currently in *Drosophila* (Morrison et al. 2000). Genes involved in RNA processing are highly conserved in *Drosophila*, so it was relatively easy to identify the 27 genes that encode small nuclear RNAs and the 99 genes that encode proteins involved in RNA processing (Mount and Salz 2000).

Annotation of the *Drosophila* genome also indicated this insect is surprisingly relevant to the study of genes and metabolic pathways involved in tumor formation and development in

humans (Potter et al. 2000). Many of the well-studied signal pathways in tumor development in humans are conserved between flies and humans, and at least 76 *Drosophila* genes that are homologous to mammalian cancer genes are under intensive study.

Furthermore, 178 (62%) of 287 known human disease genes appear to be conserved in *Drosophila*, including genes causing neurological (Alzheimer's disease, Huntington's disease, Duchenne muscular dystrophy, and juvenile-onset Parkinson's disease), renal, cardiovascular, metabolic and immune diseases, and malformation syndromes (Fortini et al. 2000). These examples of annotations and analysis represent only a small subset of the treasure trove of information that can be mined for years to come.

7.14. Bioinformatics

As the GASP project indicated, it is not always easy to find genes hidden among the thousands of nucleotide sequences produced by a genome sequencing project. Current analysis methods often only average a 70% accuracy rate in predicting structural and functional features (Bork 2000).

Part of the problem is defining a "gene" (Attwood 2000). Is a gene a heritable unit corresponding to an observable phenotype? Or is it genetic information that encodes a protein or proteins? Is it the genetic information that encodes RNA? Must the gene be translated? Is DNA a gene if the gene is not expressed? There are multiple definitions of a gene, and hence the estimates of the total number of genes in a sequenced genome can vary.

A variety of approaches have been taken to improve the process of finding genes in eukaryotes (Stormo 2000). For example, the *Drosophila* genome has **isochores**, long >300-kb DNA segments which are compositionally homogeneous on the basis of GC frequencies. As is found in humans, *Drosophila* isochores are rich in coding sequences compared to genome segments lacking high GC frequencies (Jabbari and Bernardi 2000).

The effectiveness of gene-finding programs is based on the type of information used by the program and the algorithm used to combine that information into a coherent prediction. Three types of information are used to predict the location of genes: 1) "signals in the sequence" such as splice sites; 2) "content" statistics such as codon bias; and 3) similarity to known genes (Stormo 2000).

The most important features to identify are the splice junctions of introns and exons—the donor and acceptor sites (Stormo 2000). Unfortunately, splice junctions cannot be reliably detected now, and efforts are being made to improve detection methods (Saxonov et al. 2000). Start and stop codons can be useful in predicting exons. Unfortunately, they can be uninformative if the reading frame is unknown. Some programs look for sites associated with promoters such as TATA boxes, transcription factor binding sites, and CpG islands. Poly(A) signals are used sometimes to aid in identifying the carboxyl terminus of the gene.

As the number of known coding sequences increases, the accuracy of gene-prediction programs will improve because the larger sample size of known genes will allow for more reliable statistical measures, as well as a much greater likelihood of encountering a gene that is related to one that has been identified previously.

Large genomic sequences can only be analyzed computationally, so continued improvements in analysis and annotation methods are needed (Ashburner 2000). During the past five years, advances have been made in identifying DNA sequences as coding or noncoding. Although current methods leave uncertainties, having the exact coding

prediction is unnecessary. Even partially correct predictions can focus experiments to determine the true gene structures faster than would be possible if these predictions were unavailable. Continued advances in computational and experimental methods for identifying genes, their regulatory elements, and their function are expected (Baxevanis and Ouellette 2001, Stormo 2000).

7.15. Genome Analyses of Other Arthropods

The ability to sequence genomes relatively rapidly has led to proposals to sequence the genomes of other arthropods, especially species that are of significant economic importance such as mosquitoes (especially *Anopheles gambiae*, the vector of malaria) and ticks. A proposal was made to sequence the genome of *A. gambiae* in 5 years with a cost of between $50 and $90 million (Balter 1999). In 2001, plans were finalized for sequencing the *A. gambiae* genome at a cost of approximately $10 million (Balter 2001). Preliminary analyses of the *A. gambiae* genome are already available (Favia et al. 1994, Zheng et al. 1996, Rai and Black 1999, Severson et al. 2001).

Information on genome sizes, chromosome numbers, and location of markers is available for a few insects other than *Drosophila*. For example, the genome of the silkworm *Bombyx mori* (Order: Lepidoptera) is approximately 3.8 times larger than that of *D. melanogaster*, and a substantial effort has identified approximately 400 visible mutants with approximately 200 assigned to specific linkage groups. Molecular mapping of the genome has been carried out using various DNA markers, and chromosome libraries have been constructed in BAC libraries (Shi et al. 1995, Yasukochi 1998, 1999, Wu et al. 1999).

Likewise, the flour beetle *Tribolium castaneum* (Order: Coleoptera) has been the subject of intense genetic study. *T. castaneum* has a small genome (about 200 Mb), and about 80 mutants have been located on the nine linkage groups. Linkage maps have been developed using molecular markers based on RAPD-PCR (Beeman and Brown 1999).

A linkage map of the honey bee, *Apis mellifera*, was developed based on RAPD-PCR markers (Hunt and Page 1995), and a gridded genomic library was developed for comparative and basic genetic studies (Beye et al. 1998). The honey bee is important in agriculture and is also important in behavioral research and in understanding sex determination in Hymenoptera.

Genome analyses of arthropod species other than *D. melanogaster* are highly desirable because this fly is unlikely to fully represent the entire diversity of the estimated 1 million insects.

7.16. TEs as Agents of Genome Evolution

Our perception of the role of transposable elements in genome evolution is undergoing a rapid change as more complete genomes are compared (Kidwell and Lisch 2001). TEs now may be thought of as "natural genetic engineering systems" that act to provide genetic variability and other functions (Shapiro 1999). Their designation as selfish DNA or junk DNA may be ". . . either inaccurate or misleading and . . . a more enlightened view of the transposable element–host relationship encompasses a continuum from extreme parasitism to mutualism" (Kidwell and Lisch 2001).

TEs do carry costs; they require host cell functions to replicate and proliferate, and their activity poses a risk to the host because their integration into new sites in the genome often results in deleterious mutations. In *D. melanogaster*, retrotransposons are responsible for as much as 80% of all spontaneous mutations (Miller et al. 1997).

Despite these negative aspects, TEs are abundant, and ancient components of eukaryotic genomes and their long coexistence within eukaryotic genomes suggest there has been some form of host–TE coevolution. We know that TEs can acquire a functional role in the host genome; in *D. melanogaster* the Het-A and TART retrotransposons are the telomeres (Eickbush 1997). TEs cause inversions in *Drosophila* species (Caceres et al. 1999); inversions can "tie up" specific gene combinations so that they are not scrambled during recombination, which could help to maintain favorable gene combinations.

The fact that TEs can be activated by environmental and population factors suggests that TEs could have a positive role by creating new genetic variability that could be useful under conditions which reduce the fitness of an organism (Capy et al. 2000). One hypothesis suggests that the activation of host defense genes during stress and the activation of TEs are similar processes. Alternatively, stresses could induce destabilization of the genome, leading to the malfunction of genetic systems, which would lead to the increased activity of TEs as a secondary rather than a direct effect of stress (Capy et al. 2000). Over evolutionary time, TEs have provided novel regulatory regions to preexisting host genes, and TE-derived components have undergone a molecular transition into novel host genes through a process called "molecular domestication" (Miller et al. 1997).

TEs may be more than just agents for local mutations; TEs might provide coordinated changes in the genome by inserting into a series of genes whose products already function together (Shapiro 1999). Under this scenario, different insertions could recruit new proteins into the system. During periods of extensive genome reorganization, TEs could interact with cell signals to confer on cells a far higher probability of evolving useful new multilocus systems. Thus, the relationship of TEs to their host could resemble "more symbiosis than parasitiasis" (Brosius 1999).

The association of TEs with their hosts over evolutionary time could lead to three different outcomes: 1) the coevolution of TE-derived mechanisms to minimize the negative effects of TEs on their hosts (such as transposon self-regulation, tissue specificity, targeting, and genome partitioning); 2) the evolution of host defense mechanisms, which include host suppressors; and 3) the evolution of new and altered functions of TEs in hosts (regulatory functions, structural functions, enzymatic functions, and new coding sequences) (Kidwell and Lisch 2001). Thus, TEs could enable genomes to enhance their own evolution and also serve as a major tool for generating the necessary diversity to respond to changes in the environment.

7.17. DNA Microarrays, Gene Chips, and a Laboratory-on-a-Chip

DNA microarrays or **gene chips** are revolutionizing how geneticists study gene function during development and in different tissues. DNA **microarrays** are small chips a couple of centimeters across that can be dotted with thousands of cDNA fragments from the coding regions of genes. If a sample containing fluorescently labeled RNAs that have been expressed in a cell is added to the gene chip, these labeled RNAs will anneal to the cDNA that has matching sequences through complementary base pairing. The bright fluorescent

spots will tell you which genes have been turned on. The pattern of gene expression (the **transcriptome**) offers clues as to what makes each cell, tissue, or developmental stage unique (White et al. 1999, Andrews et al. 2000).

DNA chips or microarrays also can be made with oligonucleotides that are synthesized *in situ* or by conventional synthesis, followed by on-chip immobilization. The array then is exposed to labeled sample DNA and hybridized, and complementary sequences can be determined (Ramsay 1998).

Microarray analysis is changing the way scientists conduct research because it provides a way to screen thousands of different genes in a single procedure, making it possible to do in a day what used to take years. Microarray or gene chip analysis allows us to compare the genes that are functioning during development, under different environmental conditions, and during disease.

Most microarrays are purchased from commercial sources, although it is possible to make your own (Stewart 2000, White and Burtis 2000). A *Drosophila* genome array is commercially available (from Affymetrix), and plans are underway to provide the research community with microarrays of the *Drosophila* genome by the Berkeley *Drosophila* Genome Project. Microarrays are expensive, but the cost is declining, and, relative to the amount of data that can be obtained from a single experiment, the investment may be cost-effective.

The technology of microarrays is changing rapidly (Granjeaud et al. 1999). Microarrays of genomic DNA or cDNA can be placed on nylon membranes, glass microslides, plastic, silicon, gold, a gel, or even on beads at the ends of fiber-optic bundles (Lockhart and Winzeler 2000). The technology of depositing nucleic acids on glass slides at very high densities has allowed the miniaturization of nucleic acid arrays with dramatic increases in experimental efficiency and information content. Arrays with more than 250,000 different cDNAs per square centimeter can be produced (Lockhart and Winzeler 2000).

The use of gene chips has provided some surprises and enormous amounts of data (Andrews et al. 2000). Data analysis methods are under development to interpret the results. For example, the question needs to be answered: does gene expression indicate function? At present, the conclusion is that there is a correlation between distinct expression profiles and function, but expression should not be taken as sufficient evidence for function. For example, not all genes involved in a function such as DNA replication are expressed periodically during the cell cycle, and some genes that do not need to be cell-cycle regulated are transcribed in a periodic fashion (Lockhart and Winzeler 2000).

Gene chip or microarray data require careful analysis because experimental evidence shows there is a disparity between the relative expression levels of mRNA and their corresponding proteins. Thus, expression information from both mRNA and proteins is required to understand a gene network (Dutt and Lee 2000).

Unfortunately, errors in gene chips have been found that also could lead unwary biologists to erroneous conclusions (Knight 2001). Despite the fact that microarray production is heavily automated, errors may creep in because bacterial cultures used to amplify the plasmids with the cDNAs can become contaminated. Technicians can make errors such as loading plates into the robots the wrong way around or taking samples from the wrong well for sequencing. Estimates suggest between 1 and 5% of the clones in even the best-maintained microarray sets do not contain the sequence they are supposed to. Even microarrays based on oligos can contain errors if the sequences in the databases are wrong or the wrong strand from the DNA helix is used (the noncoding strand) (Knight 2001). Other errors can occur when inadequate experimental controls are used or replications are not conducted.

Erroneous results will continue to be published until the faulty chips and experimental design methods are corrected. At the least, the sequences of the spot concerned should be verified by sequencing and by comparing the result using alternative methods of monitoring gene expression (Knight 2001, Knudsen 2002).

Gene chips almost provide "too much data." Good sample collection, data collection, and experimental design are essential to a successful experiment with microarrays. Experiments on global gene expression may yield data for thousands of genes, forcing the experimenter to consider processes, functions, and mechanisms about which we know very little. More sophisticated systems are needed to represent the data and incorporate sequence, genetics, gene expression, homology, regulation, function, and phenotype information in an organized and usable form (Lockhart and Winzeler 2000). At present, different researchers use different arrays and methods of analyzing the data, which makes it difficult to compare the results from different laboratories.

There is also an enormous effort underway to automate and miniaturize gene chips and other molecular genetic methods. The goal is to create tiny devices (perhaps 1 cm^2) that will perform the same processes that are carried out in the laboratory with large equipment. Miniaturizing could allow procedures to take place more quickly, at a lower cost, using less reagent, and with a greater resolution of detection and specificity (Talary et al. 1998). The hope is that these "laboratories-on-a-chip" will facilitate rapid advances in gene discovery, genetic mapping, and gene expression.

7.18. Proteomics: Another "-Omic"

Once we know the complete DNA sequences of organisms (genomics), the next goal is to understand how the genes are translated in living cells (proteomics). What proteins function to provide structure and function in the living organism? Proteomics was first formalized as a term in 1996 and combines "proteins" and "genomics." Definitions of proteomics and the other "-omics" are evolving, however.

Proteomics is the genome-wide analysis of proteins and includes three aspects: 1) characterization of proteins and their posttranslational modifications, 2) "differential display" to compare protein levels and types, and 3) studies of protein–protein interactions. Proteomics uses techniques such as mass spectrometry and two-dimensional protein electrophoresis (Geisow 1998, Dutt and Lee 2000, Pandey and Mann 2000).

Two-dimensional gel electrophoresis allows the identification of proteins whose expression changes in an interesting manner from that of a reference point. Two-dimensional gel electrophoresis separates proteins *by charge* using isoelectric focusing and *by size* using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Up to 11,000 proteins from a single mixture can be resolved (Dutt and Lee 2000). SDS-PAGE can purify proteins for amino acid analysis, mass spectrometry, and amino acid sequencing.

Mass spectrometry of proteins separated on two-dimensional gel electrophoresis generates different types of structural information about a protein. For example, it provides information on the mass of a protein and also generates amino acid sequence information (Dutt and Lee 2000). Furthermore, mass spectrometry can provide data on glycosylation patterns, phosphorylation, and other posttranslational modifications of proteins.

Proteomics (data on protein expression profiles) can be linked to data on nucleic acid sequence. Several software packages are available to compare multiple protein expression profiles to identify quantitative changes (Dutt and Lee 2000). For example, proteomics can

identify proteins which are associated with growth control, or with responses to high or low temperatures or to different chemicals.

Proteomics data are available in databases on the Internet, including the Protein Information Resource (PIR) (Barker et al. 2000) and SWISS-PROT (Geisow 1998, Bairoch and Apweiler 2000). These sites may contain search engines to compare sequence similarity and protein function annotations. The SWISS-PROT site provides information on the function of a protein, its domains structure, its secondary structure (alpha helix, beta sheet), its quaternary structure (homodimer, heterotrimer), similarities to other proteins, diseases associated with deficiencies in the protein, posttranslational modifications, and variants. SWISS-PROT has cross-references to additional databases (Bairoch and Apweiler 2000). The PIR site provides databases and search tools (BLAST, FASTA, pattern/peptide, pairwise alignments, multiple alignments, domain search, global or domain search, and GeneFIND) as well as technical bulletins and documentation (Barker et al. 2000).

7.19. Functional Genomics

Functional genomics has been transformed from a concept that was considered futuristic in the 1980s to an accepted part of science in the year 2002. **Functional genomics**, the assignment of function to genes, includes understanding the organization and control of genetic pathways that come together to make up the physiology of an organism (Eisenberg et al. 2000).

Using DNA microarrays or gene chips, scientists are able to analyze complex mixtures of RNA and DNA in a parallel and quantitative fashion. DNA arrays can be used to measure levels of gene expression (mRNA abundance) for tens of thousands of genes simultaneously (Brazma and Vilo 2000, Celis et al. 2000, Lockhart and Winzeler 2000). As a result of the microarray revolution, scientists are faced with an avalanche of data on mRNA expression, or, as expressed by Eisenberg et al. (2000) "piles of information but only flakes of knowledge."

7.20. Structural Genomics: Another New Horizon?

The *Drosophila* and Human Genome Projects united a large group of geneticists and others in a coordinated effort to obtain massive amounts of genomic data in a relatively short period of time. Such large-scale biology projects were unprecedented. The stage is now set for another massive undertaking by the biological community.

The new initiative is called The Structural Genomics Project (Smith 2000). Large sums of money have been allocated to the project by the United States and Japan. Once again, the project has elicited concern and apprehension among biologists because it is difficult and expensive. The Structural Genomics Project is estimated to cost more and be more complex than the Human Genome Project.

Structural genomics involves large-scale analysis of protein structures and functions based on gene sequences. Structural genomics is a new field that developed after the results of the genome sequencing projects, and recent advances in structure determination of proteins, were obtained. The Structural Genomics Project aims to link sequence, structural, and functional information and enable the prediction of unknown structures by homology modeling.

The Structural Genomics Project began in January 1998 and has moved from a concept to a well-organized, funded, consortium-based effort to determine protein structures on a large scale (Terwilliger 2000). The Structural Genomics Project aims to determine the structures of 10,000 proteins, one or more from each "fold family" within 10 years (Norvell and Machalek 2000). Once developed, this enormous body of structural data will be made available in public databases and promises to "accelerate scientific discovery in all areas of biological science" (Burley 2000).

7.21. Comparative Genomics

Now that a number of genomes have been completely sequenced, it is possible to compare genomes in order to understand the evolution of genomes (Kondrashov 1999, Rubin et al. 2000, Brown 2002).

Comparative genomics attempts to learn: how many distinct protein families are encoded in the genomes, the number of gene duplications, how similar genes are in diverse organisms, the degree of similarity of protein domains and families, and the similarity of developmental strategies and cellular processes (cell division, cell shape, cell–cell interactions) (Rubin et al. 2000). Surprisingly, approximately 30% of the predicted proteins in every organism bear no similarity to proteins in other organisms, the reasons for which remain a mystery (Rubin et al. 2000).

An example of a comparative genomic study involves a comparison of the peptidase gene families of *D. melanogaster* and *C. elegans*, which indicated differences in evolution of both form and function (Coates et al. 2000). Another comparative genomic study compared genes in the *flightless* region of the *Drosophila* genome to similar sequences in bacterial, yeast, *C. elegans*, and human genomes (Maleszka et al. 1998). Most of the 12 proteins in the *flightless* region are absent from the bacterial genomes, half are absent from yeast, but nearly all have relatives in *C. elegans* and humans, although gene order is not evolutionarily conserved.

7.22. The Post-Genomic Era: Reductionism Gives Way to Emergent Properties?

During at least the past 50 years, biology has been dominated increasingly by a reductionist approach, which narrowed the focus from the entire animal in its natural environment to increasingly smaller parts. The enormous complexity of a living organism overwhelmed the ability of biochemists, cell biologists, structural biologists, physiologists, and geneticists to study the whole animal (Vukmirovic and Tilghman 2000, White 2001). Studies of organs, then cells, and, finally, individual molecules became the focus of analyses.

The reductionist approach will continue to be productive, and necessary, to obtain detailed knowledge about gene function, gene regulation, and gene sequences of the genomes of humans, *D. melanogaster*, yeast, and *C. elegans*. This phase may be completed rather rapidly because comparisons between genomes will accelerate the process (Palsson 2000). Reductionist approaches to analysis of gene expression with microarrays will allow us to know the details of when specific genes are functioning and in what tissues.

Biology is "suddenly awash in genome-based data" and "is in the midst of an intellectual and experimental sea change" which is revolutionizing the type of questions biologists

can ask (Vukmirovic and Tilghman 2000). Leroy Hood, the leader of a team that invented the automated gene sequencer, noted recently, "The future will be the study of the genes and proteins of organisms in the context of their informational pathways or networks" (Smaglik 2000). There are so many databases available on the Web, each containing useful information on various aspects of molecular biology, that there are now databases of databases (Baxevanis 2001). Each January the journal *Nucleic Acids Research* publishes an update on the various molecular biology databases. This publication is available free on the Web.

In the near future, it should be possible to monitor simultaneously the expression of genes at the RNA or protein level, all possible protein–protein interactions, all alleles of all genes that affect a particular trait, and all protein-binding sites in a genome. It soon could be feasible to describe the properties of whole organisms in a precise and quantitative way.

Changing our focus in biology from a reductionist approach to an integrative one provides new challenges (Palsson 2000). Given the complexity of an organism, mathematical models and computer simulations probably will be required to study the integrated function of multiple gene products. Models could be required to analyze, interpret, and predict the relationship between genotype and phenotype. An early example of this integrative approach is a model that provided a comparative mathematical analysis of the genome and metabolic networks of 43 organisms representing the three domains of life (Jeong et al. 2000).

An integrative approach to biology may rely on improved bioinformatics methods and whole systems analyses to understand the properties of cellular and tissue functions and focus on the **emergent properties**, properties that arise from the whole rather than the individual parts (Palsson 2000). Future geneticists may need to enhance their computer skills and have a higher level of mathematical and informatics training. The enormous increase in knowledge of genomics during the 21st century will have at least as large an effect on the world as the changes brought about by the development of electronics and computation during the 20th century (Brent 2000).

References Cited

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, and P. G. Amanatides, et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Ambrose, B. J. B., and R. C. Pless (1987). DNA sequencing: chemical methods. *Methods Enzymol.* **152**: 522–538.

Andrews, J., G. G. Bouffard, C. Cheadle, J. Lu, K. G. Becker, and B. Oliver (2000). Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. *Genome Res.* **10**: 2030–2043.

Ashburner, M. (2000). A biologist's view of the *Drosophila* genome annotation assessment project. *Genome Res.* **10**: 392–393.

Attwood, T. K. (2000). The babel of bioinformatics. *Science* **290**: 471–472.

Bairoch, A., and R. Apweiler (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**: 45–48.

Balter, M. (1999). Gene sequencers target malaria mosquito. *Science* **285**: 508–509.

Balter, M. (2001). Sequencing set for dreaded mosquito. *Science* **291**: 1873.

Banerjee-Basu, S., J. F. Ryan, and A. D. Baxevanis (2000). The homeodomain resource: a prototype database for a large protein family. *Nucleic Acids Res.* **28**: 329–330.

Barker, W. C., J. S. Garavelli, H. Huang, P. B. McGarvey, B. C. Orcutt, G. Y. Srinivasarao, C. Xiao, L. S. L. Yeh, R. S. Ledley, J. F. Janda, F. Pfeiffer, H.W. Mewes, A. Tsugita, and C. Wu (2000). The protein information resource (PIR). *Nucleic Acids Res*. **28**: 41–44.

Barnes, W. M. (1987). Sequencing DNA with dideoxyribonucleotides as chain terminators: hints and strategies for big projects. *Methods Enzymol.* **152**: 538–556.

Baxevanis, A. D. (2001). The molecular biology database collection: an updated compilation of biological database resources. *Nucleic Acids Res.* **29**: 1–10.

Baxevanis, A. D., and B. F. Ouellette, Eds. (2001). *Bioinformatics, A Practical Guide to the Analysis of Genes and Proteins*, 2nd ed., Wiley Publ., New York.

Beeman, R. W., and S. J. Brown (1999). RAPD-based genetic linkage maps of *Tribolium castaneum*. *Genetics* **153**: 333–338.

Beye, M., A. Poch, C. Burgtorf, R. F. A. Moritz, and H. Lehrach (1998). A gridded genomic library of the honeybee (*Apis mellifera*): a reference library system for basic and comparative genetic studies of a hymenopteran genome. *Genomics* **49**: 317–320.

Bork, P. (2000). Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res.* **10**: 398–400.

Brazma, A., and J. Vilo (2000). Gene expression data analysis. *FEBS Lett.* **480**: 17–24.

Brent, R. (2000). Genomic biology. *Cell* **100**: 169–183.

Brosius, J. (1999). Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica* **107**: 209–238.

Brown T.A. (2002). *Genomes 2*. Wiley Interscience, New York.

Burley, S. K. (2000). An overview of structural genomics. *Nat. Struct. Biol. Suppl.* **10**: 932–934.

Caceres, M., J. M. Ranz, A. Barbadilla, M. Long, and A. Ruiz (1999). Generation of a widespread *Drosophila* inversion by a transposable element. *Science* **285**: 415–418.

Capy, P., G. Gasperi, C. Hiemont, and C. Bazin (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**: 101–106.

Celis, J. E., M. Kruhoffer, I. Gromova, C. Frederiksen, M. Ostergaard, T. Thykjaer, P. Gromov, J. Yu, H. Palsdottir, N. Magnusson, and T. F. Orntoft (2000). Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**: 2–16.

Cinkosky, M. J., J. W. Fickett, P. Gilna, and C. Burks (1991). Electronic data publishing and GenBank. *Science* **252**: 1273–1277.

Coates, D. R. Siviter, and R. E. Isaac (2000). Exploring the *Caenorhabditis elegans* and *Drosophila melanogaster* genomes to understand neuropeptide and peptidase function. *Trans. Biochem. Soc.* **28**: 464–469.

Davies, R. W. (1982). DNA sequencing, pp. 117–172. In: *Gel Electrophoresis of Nucleic Acids*. D. Rickwood, and B. D. Hames, Eds., IRL Press, Oxford.

Doolittle, R. F., Ed. (1990). *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Methods Enzymol. 183. Academic Press, New York.

Dutt, M. J., and K. H. Lee (2000). Proteomic analysis. *Curr. Opin. Biotechnol.* **11**: 176–179.

Eickbush. T. H. (1997). Telomerase and retrotransposons: which came first? *Science* **277**: 911–912.

Eisenberg, D., E. M. Maracotte, I. Xenarios, and T. O. Yeates (2000). Protein function in the post-genomic era. *Nature* **405**: 823–826.

Favia, G., G. Dimopoulos, and C. Louis (1994). Analysis of the *Anopheles gambiae* genome using RAPD markers. *Insect Mol. Biol.* **3**: 149–157.

FlyBase Consortium (1997). FlyBase: a *Drosophila* database. *Nucleic Acids Res.* **25**: 63–66.

Fortini, M. E., M. P. Skupski, M. S. Boguski, and I. K. Hariharan. (2000). A survey of human disease gene counterparts in the *Drosophila* genome. *J. Cell Biol.* **150**: F23–F29.

Geisow, M. J. (1998). Proteomics: one small step for a digital computer, one giant leap for humankind. *Nat. Biotechnol.* **16**: 206.

Granjeaud, S., F. Bertucci, and B. R. Jordan (1999). Expression profiling: DNA arrays in many guises. *BioEssays* **21**: 781–790.

Gribskov, M., and J. Devereux, Eds. (1991). *Sequence Analysis Primer*. Stockton Press, New York.

Hawley. R. S., and M. Y. Walker (2000). *Drosophila melanogaster*: A genome sequenced. *Genome Res.* **10**: 593–594.

Howe, C. J., and E. S. Ward, Eds. (1989). *Nucleic Acids Sequencing. A Practical Approach*. IRL Press, New York.

Hunkapiller, T. R., J. Kaiser, B. F. Koop, and L. Hood (1991). Large-scale and automated DNA sequence determination. *Science* **254**: 59–67.

Hunt, G. J., and R. E. Page (1995). Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics* **139**: 1371–1382.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Jabbari, K., and G. Bernardi (2000). The distribution of genes in the *Drosophila* genome. *Gene* **247**: 287–292.

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. (2000). The large-scale organization of metabolic networks. *Nature* **407**: 651–654.

Jeyaprakash, A., and M. A. Hoy (1995). Complete sequence of a *mariner* transposable element from the predatory mite *Metaseiulus occidentalis* isolated by an inverse PCR approach. *Insect Mol. Biol.* **4**: 31–39.

Kafatos, F. C., C. Louis, C. Savakis, D. M. Glover, M. Ashburner, A. J. Link, I. Siden-Kiamos, and R. D. C. Saunders (1991). Integrated maps of the *Drosophila* genome: progress and prospects. *Trends Genet.* **7**: 155–161.

Kidwell, M. G., and D. R. Lisch (2001). Transposable elements, parasitic DNA and genome evolution. *Evolution* **55**: 1–24.

Knight, J. (2001). When the chips are down. *Nature* **410**: 860–861.

Knudsen, S. (2002). *A Biologist's Guide to Analysis of DNA Microarray Data*. Wiley Interscience, New York.

Kondrashov, A. S. (1999). Comparative genomics and evolutionary biology. *Curr. Opin. Genet. Dev.* **9**: 624–629.

Koshland, D. E. (1991). Flying into the future. *Science* **254**: 173.

Lanave, C., S. Liuni, F. Licciulli, and M. Attimonelli (2000). Update of the AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Res.* **28**: 153–154.

Lasko, P. (2000). The *Drosophila melanogaster* genome: translation factors and RNA binding proteins. *J. Cell Biol.* **150**: F51–F56.

Lockhart, D. J., and E. A. Winzeler (2000). Genomics, gene expression and DNA arrays. *Nature* **405**: 827–836.

Maleszka, R., H. G. De Couet, and G. L. G. Miklos (1998). Data transferability from model organisms to human beings: insights from the functional genomics of the *flightless* region of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **95**: 3731–3736.

Maxam, A. M., and W. Gilbert (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**: 560–564.

Maxam, A. M., and W. Gilbert (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**: 499–560.

Meldrum, D. (2000a). Automation for genomics, part one: Preparation for sequencing. *Genome Res.* **10**: 1081–1092.

Meldrum, D. (2000b). Automation for genomics, part two: Sequencers, microarrays and future trends. *Genome Res.* **10**: 1288–1303.

Merriam, J. M. Ashburner, D. L. Hartl, and F. C. Kafatos (1991). Toward cloning and mapping the genome of *Drosophila*. *Science* **254**: 221–225.

Mierendorf, R. C., and D. Pfeffer (1987). Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* **152**: 556–562.

Miller, W. J., J. F. McDonald, and W. Pinsker (1997). Molecular domestication of mobile elements. *Genetica* **100**: 261–270.

Misra, S., M. A. Crosby, and R. A. Drysdale (2000). Using *Drosophila* genome databases,. pp. 509–523. In: *Drosophila Protocols*. W. Sullivan, M. Ashburner, and R. S. Hawley, Eds. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Morrison, D. K., M. S. Murakami, and V. Cleghon (2000). Protein kinases and phosphatases in the *Drosophila* genome. *J. Cell Biol.* **150**: F57–F62.

Mount, D. W. (2001). *Bioinformatics. Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mount, S. M., and H. K. Salz (2000). Pre-messenger RNA processing factors in the *Drosophila* genome. *J. Cell Biol.* **150**: F37–F43.

Norvell, J. C., and A. Z. Machalek (2000). Structural genomics programs at the US National Institute of General Medical Sciences. *Nat. Struct. Biol. Suppl.* **10**: 931.

Palsson, B. (2000). The challenges of in silico biology. *Nat. Biotechnol.* **18**: 1147–1150.

Pandey, A., and M. Mann (2000). Proteomics to study genes and genomes. *Nature* **405**: 837–846.

Pennisi, E. (2000a). Fruit fly genome yields data and a validation. *Science* **287**: 1374.

Pennisi, E. (2000b). Ideas fly at gene-finding jamboree. *Science* **287**: 2182–2184.

Perier, R. C., V. Praz, T. Junier, C. Bonnard, and P. Bucher (2000.) The eukaryotic promoter database (EPD). *Nucleic Acids Res.* **28**: 302–303.

Potter, C. J., G. S. Turenchalk, and T. Xu (2000). *Drosophila* in cancer research. *Trends Genet.* **16**: 33–39.

Rai, K. S., and W. C. Black, IV. (1999). Mosquito genomes: structure, organization and evolution. *Adv. Genet.* **41**: 1–33.

Ramsay, G. (1998). DNA chips: state-of-the art. *Nat. Biotechnol.* **16**: 40–44.

Reese, M. G., G. Hartzell, N. L. Harris, U. Ohler, J. F. Abril, and S. E. Lewis (2000). Genome annotation assessment in *Drosophila melanogaster*. *Genome Res.* **10**: 483–501.

Roberts, R. J., and D. Macelis (2000). REBASE—restriction enzymes and methylases. *Nucleic Acids Res.* **28**: 306–307.

Rubin, G. M., and E. B. Lewis (2000). A brief history of *Drosophila*'s contributions to genome research. *Science* **287**: 2216–2218.

Rubin, G. M., M. D. Yandell, J. R. Wortman, G. L. G. Kiklos, C. R. Nelson, I. K. Hariharan, M. E. Fortini, et al. (2000). Comparative genomics of the eukaryotes. *Science* **287**: 2204–2215.

Sambrook, J., and D. W. Russell (2001). DNA sequencing. Chapter 12. In: *Molecular Cloning, A Laboratory Manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sambrook, J., E. F. Fritsch, T. Maniatis (1989). DNA sequencing. Chapter 13. In: *Molecular Cloning, A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., S. Nicklen, and A. R. Coulson (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5463–5467.

Saxonov, S. I. Daizadeh, A. Fedorov, and W. Gilbert (2000). EID: the exon–intron database—an exhaustive database of protein-coding intron-containing genes. *Nucleic Acids Res.* **28**: 185–190.

Schisler, N. J., and J. D. Palmer (2000). The IDB and IEDB: intron sequence and evolution databases. *Nucleic Acids Res.* **28**: 181–184.

Sealey, P. G., and E. M. Southern (1982). Electrophoresis of DNA, pp. 39–76. In: *Gel Electrophoresis of Nucleic Acids*. D. Rickwood and B. D. Hames, Eds. IRL Press, Oxford.

Severson, D. W., S. E. Brown, and D. L. Knudson (2001). Genetic and physical mapping in mosquitoes: Molecular approaches. *Annu. Rev. Entomol.* **46**: 183–219.

Shapiro, J. A. (1999). Transposable elements as the key to a 21st century view of evolution. *Genetica* **107**: 171–179.

Shi, J., D. G. Heckel, and M. R. Goldsmith (1995). A genetic linkage map for the domesticated silkworm, *Bombyx mori*, based on restriction fragment length polymorphisms. *Genet. Res.* **66**: 109–126.

Smaglik, P. (2000). *For my next trick. . .Nature* **407**: 828–829.

Smith, T. (2000). *A new era. Nat. Struct. Biol. Suppl.* **10**: 927.

Stein, L. (2001). Genome annotation. *Nat. Rev. Genet.* **2**: 493–503.

Stewart, D. J. (2000). Making and using DNA microarrays: a short course at Cold Spring Harbor Laboratory. *Genome Res.* **10**: 1–3.

Stormo, G. D. (2000). Gene-finding approaches for eukaryotes. *Genome Res.* **10**: 394–397.

Tabor, S., and C. C. Richardson (1987). DNA sequence analyses with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84**: 4767–4771.

Talary, M. S., J. P. H. Burt, and R. Pethig (1998). Future trends in diagnosis using laboratory-on-a-chip technologies. *Parasitology* **117**: S191–S203.

Terwilliger, T. C. (2000). Structural genomics in North America. *Nat. Struct. Biol. Suppl.* **10**: 935–939.

Venter, J. C., H. O. Smith, and L. Hood (1996). A new strategy for genome sequencing. *Nature* **381**: 364–366.

Venter, J. C., et al. (2001). The sequence of the human genome. *Science* **291**: 1304–1351.

Vukmirovic, O. G., and S. M. Tilghman (2000). Exploring genome space. *Nature* **405**: 820–822.

Weir, B. S. (1990). *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.

Wheeler, D. L., C. Chappey, A. E. Lash, D. D. Leipe, T. L. Madden, G. D. Schuler, T. A. Tatusova, and B. A. Rapp (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **28**: 10–14.

White, K. P. (2001). Functional genomics and the study of development, variation and evolution. *Nature Rev. Genet.* **2**: 528–537.

White, K. P., and K. C. Burtis (2000). *Drosophila* microarrays: from arrayer construction to hybridization, pp. 487–507. In: *Drosophila Protocols*. W. Sullivan, M. Ashburner, and R. S. Hawley, Eds. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

White, K. P., S. A. Rifkin, P. Hurban, and D. S. Hogness (1999). Microarray analysis of *Drosophila* development during metamorphosis. *Science* **286**: 2179–2184.

Wu, C., S. Asakawa, N. Shimizu, S. Kawasaki, Y. Yasukochi (1999). Construction and characterization of bacterial artificial chromosome libraries from the silkworm, *Bombyx mori*. *Mol. Gen. Genet.* **261**: 698–706.

Yasukochi, Y. (1998). A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. *Genetics* **150**: 1513–1525.

Yasukochi, Y. (1999). A simple and accurate method for generating co-dominant markers: an application of conformation-sensitive gel electrophoresis to linkage analysis in the silkworm. *Mol. Gen. Genet.* **261**: 796–802.

Zheng, L., M. Q. Benedict, A. J. Cornel, F. H. Collins, and F. C. Kafatos (1996). An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics* **143**: 941–952.

Some Relevant Web Sites

Brown Lab Homepage (cmgm.stanford.edu/pbrown)
Interactive Fly (sdb.bio.purdue.edu/fly/aimain/1aahome.htm)
FlyBase (flybase.bio.indiana.edu)
Drosophila Virtual Library (ceolas.org/fly/)
SWISS-PROT protein sequence database (http://www.expasy.ch/sprot/) and (http://www.ebi.ac.uk/swissprot/)
Protein Information Resource, PIR (http://pir.georgetown.edu) and (http://www.mips.biochem.mpg.de)
Gene-Chips home page provides information about DNA microarrays (www.gene-chips.com)
Nature Genome Gateway (http://www.nature.com/genomics/gateway)
Institute for Genomic Research (www.tigr.org)

8

DNA Amplification by the Polymerase Chain Reaction: Molecular Biology Made Accessible

8.1. Overview

Occasionally a technique is developed that changes the kinds of questions that can be answered in biology. Within a few short years, the polymerase chain reaction (PCR) became just such a powerful tool for solving a myriad of basic and applied problems. Modifications of the PCR continue to be developed, and these permit additional applications.

The PCR is a method for amplifying (copying) small amounts of DNA or RNA. It can be used to isolate specific DNA fragments, end label DNA, clone cDNA and genomic DNA, sequence DNA, mutate specific DNA sequences, alter promoters, quantitate the amount of RNA or DNA, and identify molecular markers for taxonomic or ecological studies. The PCR requires a DNA polymerase, dNTPs, template DNA, and primers. Information about sequences at each end of the DNA to be amplified is needed in order to synthesize appropriate primers for "standard" PCR. When two specific primers are used, amplification of DNA theoretically is geometric, producing large quantities of specific DNA suitable for sequencing, cloning, or probing. PCR methods that use single primers, such as Random Amplified Polymorphic DNA-PCR (RAPD-PCR), also can result in DNA amplification. RAPD-PCR uses short, randomly chosen primers to amplify multiple DNA segments in the genome. The resulting banding patterns (similar to bar codes) provide information about genetic variation within the entire genome of insects.

The power of the PCR to amplify DNA is dramatic; theoretically even a single molecule can be amplified, although efficiency usually is lower. This power creates formidable problems with contamination and requires careful organization of PCR experiments and the use of adequate controls. The relative ease with which the PCR can be used by novices in molecular biology has made it possible for a diverse group of biologists to use the PCR to study molecular systematics, evolution, ecology, behavior, and development.

8.2. Introduction

The **polymerase chain reaction (PCR)** is an *in vitro* or cell-free method for synthesizing DNA or RNA sequences in nearly any amount required, starting with a small initial quantity. The PCR is one of the most accessible and versatile techniques available to entomologists interested in both basic and applied problems. The PCR is powerful because it can be used to isolate specific DNA fragments, end label DNA, mutagenize specific DNA fragments, clone cDNA and genomic DNA, sequence DNA, quantitate RNA and DNA, and alter a variety of sequences to study gene expression.

DNA polymerase, as used in the PCR, was designated the "Molecule of the Year" by *Science* in 1989 (Guyer and Koshland 1989). The PCR became a standard laboratory method in an extraordinarily short time after it was invented in 1985 (Mullis 1987, Rabinow 1996). In 1993, Kary Mullis received the Nobel Prize for chemistry for his work on the PCR, although some believe other scientists should have shared credit for the invention (Rabinow 1996).

Improvements in, and optimization of, the PCR since 1985 have led to its use by numerous scientists. The PCR has become a common procedure in forensics and diagnostics and is revolutionizing studies of basic biology, ecology, and evolution. This relatively simple technique has provided scientists with limited experience in molecular biology the opportunity to apply molecular techniques to diverse problems (White et al. 1989, Arnheim et al. 1990, Erlich and Arnheim 1992).

Although the PCR is conceptually simple, the process is, in fact, not completely understood. The PCR involves complex kinetic interactions between the template (or target) DNA, product DNA, oligonucleotide primers (polymers of 10 to ~30 nucleotides), deoxynucleotide triphosphates (dNTPs), buffer, and enzyme (one or more DNA polymerases). These relationships change during the course of the reaction (Figure 8.1).

The PCR works well with most DNA targets, but adjustments usually are needed in the reaction parameters in order to improve specificity and yield. A number of parameters can be modified to optimize the PCR, including: the reaction buffer (particularly the $MgCl_2$ concentration); relative concentrations of template DNA, primers, dNTPs, and DNA polymerase; annealing time and temperature; and extension time and temperature (Carbonari et al. 1993). No single protocol is appropriate for all situations, and each new experiment requires optimization. For example, amplifying a 100-bp fragment is not equivalent to amplifying a 10-kb DNA fragment. Modifying each of the components of the PCR to develop an optimized reaction is time consuming and tedious, but kits now are available that allow one to optimize the PCR with fewer steps.

This chapter describes what we know about the basic PCR, discusses some of the modifications of the basic method, identifies applications of the PCR, and provides

references to additional information. PCR technology is changing rapidly, and new applications and methods of significance to entomologists will no doubt continue to become available.

8.3. The Basic Polymerase Chain Reaction (PCR)

The PCR involves combining a DNA sample (the template) with oligonucleotide primers, deoxynucleotide triphosphates (dNTPs), and a DNA polymerase in a buffer. The specificity of the basic PCR depends on base pairing by the two primers to the template (target) DNA (Figure 8.1A).

The **primers**, which are single-stranded (ss) sequences that flank the DNA to be amplified, anneal to the single strands of template DNA that has been denatured by heating it. Repeated PCR cycles involve heat denaturation to separate the template DNA strands, cooling to allow annealing of primers to the complementary DNA sequences of the ss template DNA, and "extension" (or replication) of new (product) DNA strands by DNA polymerase. The base sequence of the new strand is determined by the sequence of the ss template DNA. DNA synthesis proceeds across the region *between* the annealed primers (Figure 8.1B). This mixture is repetitively heated to separate the ds DNA and cooled until the desired amount of template DNA has been amplified.

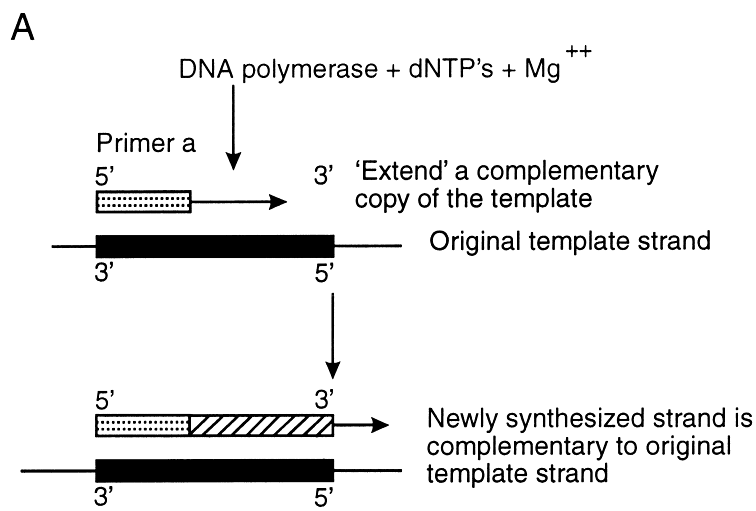


Figure 8.1. The standard, allele-specific PCR protocol. A) Template DNA is isolated and mixed with primers, dNTPs, and *Taq* DNA polymerase in a buffer with Mg^{2+}. The double-stranded template DNA is heated to denature it so that the primer can anneal to single-stranded target DNA (only one template strand is shown). *Taq* synthesizes a new single strand of complementary DNA using the primer to initiate synthesis. The dNTPs are added in a sequence determined by the template DNA strand. This initial extension continues on beyond the desired end, as shown by the arrow on the newly synthesized strand in the lower diagram. The process of denaturation, annealing of the primer to the template, and DNA synthesis (or extension) is called a cycle.

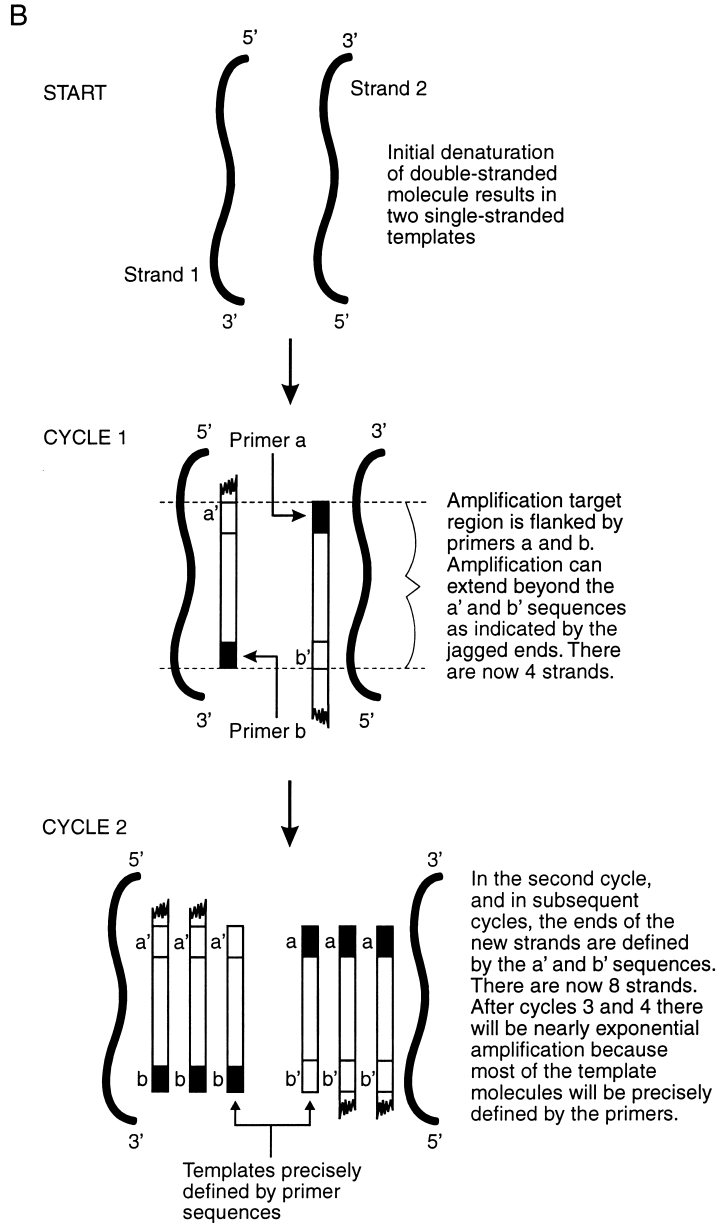


Figure 8.1. B) The process of amplification has low efficiency in the early cycles, and some of the products produced also lack defined ends. At the start a single molecule (2 strands) is denatured by heating, and primer a and b, respectively, anneal to homologous sequences on strand 1 and 2. As shown above, in cycle 1 amplification extends some distance along the original DNA template strands. During cycle 2, the DNA is once again heated to denature it, and old and new strands can serve as templates for DNA synthesis. Note, the DNA is synthesized from the 5′ to the 3′ direction. The primers are indicated by the small rectangles; the sequences homologous to the primers are designated at a′ and b′. Once again in cycle 2, extension of some new strands occurs beyond the a′ or b′ sequences. However, for the first time templates are produced with precisely defined ends. After the next couple of cycles (cycles 3 and 4), most of the templates will be precisely defined, and it is only then that amplification becomes nearly geometric.

8.3.1. The First Few Cycles Are Critical

All cycles begin by denaturing the template DNA (and any previously synthesized product) so that the template DNA and newly synthesized product DNA become single-stranded (Figure 8.1A). As the temperature is lowered, the primers anneal to the complementary sequences of the DNA. The annealing step in the early cycles requires the primers to "scan" the template DNA for the correct target sequences to which they can anneal (Ruano et al. 1991, Dieffenbach 1995, Harris and Jones 1997). Because much of the template DNA will not have the correct complementary sequence, annealing during the early cycles may not be as efficient as it is during the middle cycles. Improper interactions of primers with template in the first few cycles can lead to a nonspecific product. The PCR product will be specific *only* if the two primers bind to sites on either complementary strand of the DNA and, for the standard PCR, these sites are not more than about 1 to 2 kb apart. Thus, the first few cycles are very important if accurate and high yields are to be produced.

During the middle cycles, the DNA product previously synthesized is the preferred template for the primers, so the target template now is perfectly demarcated (Figure 8.1). Finally, in the late cycles, denatured amplified products that are present in high concentration can hybridize to themselves, thus blocking the primers from their complementary sites. In theory, DNA sequences up to about 10 kb in length can be synthesized by the standard protocol, but sequences of 2 kb, or less, are more readily obtained.

8.3.2. PCR Power

The power of the PCR is based on the fact that the products of one replication cycle serve as a template for the next. Each successive cycle, in theory, doubles the number of DNA molecules synthesized in the previous cycle, resulting in the exponential accumulation of the target DNA at approximately 2^n, where n is the number of cycles. In practice, the PCR is never 100% efficient, and less product will be produced. The early cycles are less efficient than the middle cycles because precisely defined template strands predominate only after the first few cycles (Figure 8.1B). Late in the PCR, the availability of various components may limit the yield, including primer concentration, dNTPs, or DNA polymerase (Czerny 1996).

8.3.3. Standard PCR Protocols

Table 8.1 gives a procedure suitable for amplifying genomic DNA from *Drosophila* and demonstrates the relative simplicity of the technique. Table 8.2 discusses some of the issues that must be considered in setting up new PCRs.

The PCR is performed using commercially available temperature cyclers that allow the programming of the three fundamental reaction temperatures during **denaturing**, **annealing**, and **extension** (Figure 8.2). A typical amplification cycle involves denaturing the template DNA at 94°C for 20 sec, annealing the primers to the template at 55°C for 20 sec, and extending (or synthesizing) the DNA at 72°C for 30 sec. Because the instruments require time to heat and cool to a specific temperature, each actual cycle time may require 10 min or more, depending upon the specific machine used. If 25 cycles are performed, the total time will be approximately 4 h and the target DNA will have

Table 8.1. Example of a Standard Allele-Specific PCR Reaction Protocol for Amplifying *Drosophila* DNA

1. Set up a 100-µl reaction in a 0.5-ml microfuge tube, mix, and overlay with 75 µl of mineral oil:
   - Template DNA (10^5 to 10^6 target molecules)
   - 20 pmol each primer (each primer 18 to 30 nucleotides long)
   - 100 mM Tris-HCl (pH 8.3 at 20°C)10 mM $MgCl_2$
   - 0.05% Tween 20
   - 50 µM each dNTP
   - 2 units of *Taq* DNA polymerase
2. Perform 25 to 35 cycles of PCR using the following temperature profile:

Denaturation	96°C, 15 seconds (a longer initial time is desirable)
Primer annealing	55°C, 30 seconds
Primer extension	72°C, 1.5 minutes

3. Cycling should conclude with a final extension at 72°C for 5 minutes. Stop reactions by chilling to 4°C and/or adding EDTA (ethylene dinitrilotetraacetic acid, a chelating agent) to 10 mM.

been amplified approximately one millionfold, assuming a doubling has occurred in each cycle.

8.3.4. DNA Polymerases

The PCR, as first described in 1985, used the **Klenow fragment** of *Escherichia coli* DNA polymerase I to produce copies of target DNA (Mullis and Faloona 1987, Saiki et al. 1985). Because the Klenow fragment is heat sensitive, fresh enzyme had to be added to each cycle, making the PCR an exceedingly tedious procedure!

The efficiency and fidelity of the PCR were dramatically improved by employing a heat-resistant polymerase (***Taq* DNA polymerase**) so that the procedure could be carried out at high temperatures, yet the enzyme did not have to be added before each cycle (Saiki et al. 1988, Eckert and Kunkel 1991, Taylor 1991, Goodman 1995). *Taq* was isolated from the bacterium *Thermus aquaticus*, which was collected from a hot spring in Yellowstone National Park. Because *Taq* can withstand repeated exposures to temperatures up to 94°C, its use greatly increased the ease with which the PCR could be performed.

Taq DNA polymerase is a 94-kDa protein with a temperature optimum of approximately 75° to 85°C. It can extend (add on) more than 60 nucleotides per sec at 70°C with a GC-rich 30-mer primer. In a PCR mixture, *Taq* retains 50% of its activity after about 40 min at 94°C. The use of *Taq* thus increased the specificity and yield of the PCR over that possible with the Klenow fragment because the primers could be annealed and extended at higher temperatures, which eliminated much of the *nonspecific* amplification. Longer PCR products were produced because the secondary structure of the template DNA was eliminated at these higher temperatures, as well. Fragments about 500 bp long can be synthesized with the Klenow fragment, but fragments up to 10 kb long sometimes can be produced with *Taq*.

8.3.5. Other Thermostable DNA Polymerases

Genetically engineered variants of *Taq* have been developed, and DNA polymerases from other sources now are available commercially in native and cloned form (Engelke et al. 1990,

Table 8.2. Optimizing Standard PCR Reactions Involves Optimizing Reaction Components

PCR component	Issues to consider
Primer	1. Select primers with a random base distribution and GC content similar to template DNA being amplified. 2. Avoid primers with stretches of polypurines and polypyrimidines or other unusual sequences. 3. Avoid sequences with a secondary structure, especially at the 3′ end. 4. Check primers for complementarity and avoid primers with 3′ overlaps to reduce primer–dimer artifacts. 5. Construct primers 20 to 30 nucleotides long. 6. Optimize the amount of primers used. 7. Design so the base at the 3′ end of the primer is a G or C to enhance specificity (G-C clamp).
Template DNA	1. Template DNA should be free of proteases that could degrade the DNA polymerase. 2. Template DNA with high levels of proteins or salts should be diluted or cleaned up to reduce inhibition of DNA polymerase activity. 3. Highly concentrated template DNA may yield nonspecific product or inhibit the reaction. 4. It is rare that template DNA concentration is too low.
PCR buffer	1. $MgCl_2$ concentration is very important. 2. Excess Mg^{2+} promotes production of nonspecific product and primer-dimer artifacts. 3. Insufficient Mg^{2+} reduces yields. 4. Presence of EDTA or other chelators can reduce the availability of Mg^{2+}.
Taq polymerase	1. Excessive *Taq* concentrations can yield nonspecific products and reduce yield. Recommended concentrations are between 0.5 and 2.5 units per 100 µl reaction. Add the *Taq* at 94°C and mix thoroughly. 2. Stringency can be increased by increasing the annealing temperature, adjusting dNTP concentrations, and minimizing incubation time.
dNTPs	1. dNTP concentrations should be equivalent to minimize misincorporation errors. 2. Low dNTP concentrations minimize mispriming, but if too low, can reduce the amount of product.
Cycle parameters	
Incubation	1. Time varies with length of target being amplified; 1 min/kb is average. 2. Ramp time (time to change from one temperature to another) should be minimized to improve specificity. 3. Insufficient step is a common problem; 94°C results in complete separation, but excess time can cause denaturation of *Taq* polymerase.
Annealing	1. Annealing temperature depends on length and GC content of primers; 55°C good for primers 20 nt long (50% GC). 2. Higher annealing temperatures may be needed to increase primer specificity.
Cycle number	1. Optimum number varies with starting concentration of template DNA, and all of the previously listed parameters. 2. Too many cycles increases amount of nonspecific product; too few results in a low yield that can't be detected by gel electrophoresis. 3. If additional product is required, it is better to reamplify, using an aliquot of the first reaction as the template, than to increase the number of cycles.

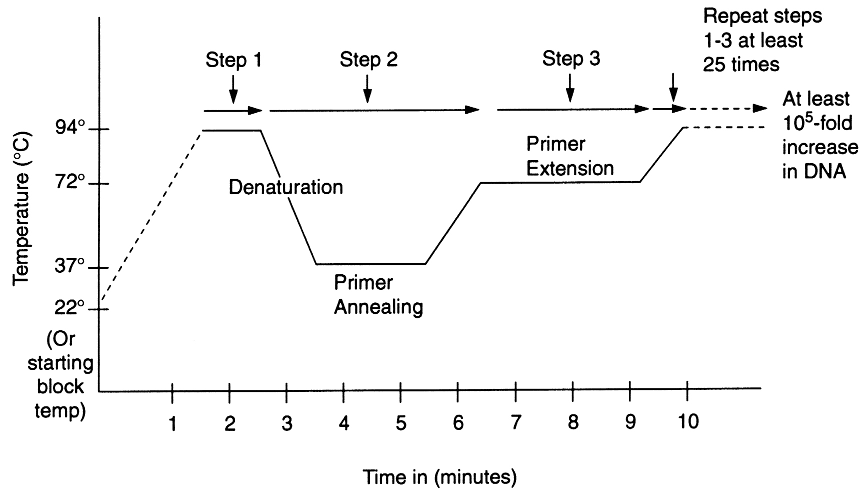


Figure 8.2. Example of a typical PCR protocol. Step 1 involves denaturing the double-stranded DNA template at 94°C. Step 2 involves annealing the primer to the single-stranded target DNA by base pairing. Step 3 involves synthesis of new DNA from the 3′ end of the primer (= primer extension) by DNA polymerase using dNTPs in a sequence determined by the template DNA. Steps 1 through 3 are a cycle, and approximately 25 cycles will yield an increase in DNA content by a factor of approximately 1 million (2^{25}).

Erlich et al. 1991, Perler et al. 1996). For example, a recombinant *T. aquaticus* polymerase called the "Stoffel fragment" persists at 97.5°C and exhibits optimal activity over a broad range of Mg^{2+} concentrations.

Other thermostable DNA polymerases have been isolated from Eubacteria (*Thermus* and *Bacillus*) and Archaea (*Thermococcus*, *Pyrococcus*, and *Sulfolobus*). Many are commercially available and have specific attributes such as 3′ to 5′ exonuclease activity, "proofreading" ability, different molecular weights, and different stabilities and temperature optima (Perler et al. 1996). For example, a thermostable enzyme isolated from *T. thermophilus* can reverse-transcribe RNA efficiently at high temperatures. The thermostability of this enzyme appears to minimize the importance of secondary structure in the RNA template and allow efficient cDNA synthesis at high temperatures.

A DNA polymerase isolated from the archaebacterium *S. acidocaldarius* carries out polymerization at 100°C, which could facilitate amplification of DNA regions with secondary structure (Arnheim and Erlich 1992). Polymerases from *Thermoplasma acidophilum*, *Thermococcus litoralis*, and *Methanobacterium thermoautotrophicum* have 3′ to 5′ exonuclease activities, which means that they can proofread, reducing the rate of misincorporation or errors.

8.3.6. Primers Are Primary

Although all the components of a PCR are important, primers are truly crucial (Table 8.2). Well-designed primers can result in 100- to 1000-fold increases in sensitivity (He et al. 1994a,b). As a general rule, longer primers are better than shorter for increasing specificity,

but length is not the only consideration. Many of the modifications of the PCR have involved modifying the number, size, and specificity of the primers used, as described below.

What is a primer and where do you get them? A primer is a short (10 to ∼30 nucleotides) single-stranded polymer of oligonucleotides. The standard PCR (allele-specific) PCR requires that the specific sequence of the DNA targeted for amplification be known in order to synthesize a primer. Thus, information is required about the gene/DNA to be amplified. Primers anneal to the target DNA by complementary base pairing, with A annealing to T, and C to G. The primers determine the length, specificity, and nature of the DNA fragment amplified.

Allele-specific (standard) PCR requires a pair of primers that flank the target DNA to be amplified; extension (copying of the single strand of template DNA) occurs from each 3′OH end of the primer, so that the ends of the amplified DNA are defined by the 5′ ends of the primers. As a result, the length of the DNA generated is equal to the lengths of the two primers plus the length of the template DNA (Figure 8.1).

Most primers are synthesized to order on a DNA synthesizer. A number of commercial suppliers provide this service, with the price determined by the number of bases in the primers. Primers may be called 10-, 20-, or 30-mers, based on their length. Primers can be constructed that contain extensions so that restriction enzyme sites, regulatory codons, or labels can be added to the target DNA. These sequences will be incorporated into the 5′ end of the target sequence, making the products easier to clone or sequence.

Selecting primers for allele-specific PCR remains somewhat empirical, although computer programs have been developed to aid in their design (Table 8.2) (Lexa et al. 2001). It is desirable, where possible, to select primer pairs with a G+C content of around 50% and a random base distribution (except at the 3′ end). It is important to avoid complementary 3′ ends of the primer pairs to avoid primer-dimer artifacts that will reduce the yield of the desired DNA.

Runs of three or more Cs or Gs at the 3′ ends of primers may promote mispriming at G+C-rich regions. Primers with T, C, or G as the 3′ (last) nucleotide result in a more specific PCR product than if the primer ends in an A (Ayyadevara et al. 2000). Amplification efficiency is reduced when T and A occupy the penultimate 3′ position of the primer (Ayyadevara et al. 2000). Palindromic sequences within primers should be avoided, as should sequences that will yield a significant secondary structure. In some cases, primers with two Gs and/or Cs at the 3′ end (the G/C clamp) will ensure the primer anneals strongly to the template to promote specific priming (Roux 1995).

Sometimes suboptimal primers, perhaps containing high amounts of A and T, must be used because of the nature of the target sequence. Low concentrations of tetramethylammonium chloride (TMAC) could reduce mispriming and thus reduce nonspecific amplification (Chevet et al. 1995).

8.3.7. Storing Insects for the PCR

Ideal killing and storage techniques include placing the insects into an ultralow freezer (–80°C) or into liquid nitrogen or dry ice. Rapid killing reduces damage to DNA by endogenous DNases. Storage of insects under inappropriate conditions can have detrimental effects on the quality and quantity of DNA available for the PCR (Dick et al. 1993). However, it is not always possible to kill and store insects under optimal conditions in remote field conditions.

Alternative killing and storage methods include the use of ethanol (EtOH) at 95 or 100% (Quicke et al. 1999). The use of EtOH at less than 95% is undesirable because the water in insects dilutes the EtOH, which can result in degradation of DNA. If your insects are large, it may be desirable to kill them in 100% EtOH, pour it off, and replace it with fresh to reduce the dilution with endogenous water.

Other killing and storage methods may provide useful DNA, although loss in quality and quantity may occur. Desiccation with silica gel has preserved DNA of tiger beetles for several months (Vogler and Pearson 1996). Other preservation methods include acetone, 2-propanol, diethyl ether and ethyl acetate, which allows insects to be stored for about six months (Fukatsu 1999). Fukatsu (1999) recommended the use of acetone which preserved the DNA of insects (as well as the DNA of microbial organisms within the insects) for more than two years at room temperature.

Many small insects are processed by critical point drying or other drying techniques for preservation in museums (Austin and Dillon 1997). Such dried insects sometimes can be used for molecular studies, especially if the target DNA is short and abundant (such as mitochondrial and ribosomal DNA). Specimens that were killed in 100% EtOH, stored at 5°C, and then dried yielded good-quality DNA upon extraction (Austin and Dillon 1997). Amplification of long segments of single-copy genes from insects that have been poorly preserved is likely to be difficult.

Storage in methanol and chloroform, as well as low concentrations of EtOH, can result in poor preservation of DNA (Fukatsu 1999).

8.3.8. Preparing DNA Samples

Template DNA used in the PCR generally should be free of proteases that could degrade the DNA polymerase. It should be free of nucleases that could degrade DNA, and free of DNA binding proteins or high levels of heat-precipitable proteins that would inhibit amplification. Ideally, 10^5 to 10^6 template DNA molecules will be available, although successful PCRs have been achieved with only a few DNA molecules.

Relatively crude DNA preparations *sometimes* can be used for the PCR. For example, when large numbers of individual insects must be processed, it is possible to do PCR on undissected larval or adult insects without prior isolation of the DNA (Grevelding et al. 1996). Apparently, the repeated denaturation steps at high temperature are sufficient to lyse cells so that sufficient template DNA is available, especially when the target DNA is present in high copy number in each cell (such as mitochondrial genes or ribosomal RNA genes). The advantage of using a crude lysate is that it reduces the time and costs to prepare the sample, which is important when hundreds or thousands of specimens must be evaluated. Such crude preparations do not allow the DNA to be stored, however.

Another crude preparation method involves boiling the insect with subsequent dilution. Lysing cells in boiling water is a quick and effective method of preparing DNA for PCR, although only a small volume of the extract can be used because cellular debris may inhibit the PCR.

Cells in complex biological fluids or cells resistant to lysis require additional processing. Rapid and inexpensive DNA extraction can be achieved using Chelex, a polyvalent chelating agent in resin form, which reduces degradation of DNA heated in low-ionic-strength buffers, probably by chelating heavy-metal ions that may serve as catalysts in the breakdown of DNA (Singer-Sam et al. 1989). Adding Chelex during boiling appears to increase the amount of DNA produced from samples containing small amounts of template. Chelex is nontoxic,

provides rapid results, and can be used to isolate DNA from hundreds of individual insects or mites suitable for either the standard PCR or RAPD-PCR (Edwards and Hoy 1994). A disadvantage to Chelex is that the extracted DNA is unstable and must be used within a few days. Other rapid extraction methods are available commercially and may produce repeatable results. A variety of DNA extraction methods have been tested for the PCR (Steiner et al. 1995, Goldenberger et al. 1995, Hammond et al. 1996, Shahjahan et al. 1995, Aljanabi and Martinez 1997), and many commercial kits are available. However, these may be expensive if used to extract DNA from large numbers of samples.

The "best" DNA extraction method will depend on the goals of your experiment. Preliminary experiments should be conducted to determine which DNA extraction method is appropriate. For example, if the goal is to process large numbers of insects for ecological studies, then rapid, nontoxic, and inexpensive extraction methods, such as Chelex, may be useful and sufficient, especially if amplifying multiple copy genes in mitochondria or ribosomes. However, Chelex may yield some false negative results and the extracted DNA cannot be stored. If it is important to maintain live insects for further study, it is possible to extract insect hemolymph from large insects, extract the DNA from the hemolymph with Chelex, and yield DNA suitable for microsatellite analysis (Gerken et al. 1998).

Some experiments require specialized extraction methods (Mauel et al. 1999). For example, different numbers of false negative results were obtained when DNA from ticks infected with a pathogen, granulocytic ehrlichiosis, was extracted by three different methods. Blood-fed ticks have inhibitors of the PCR that cannot be extracted easily with standard extraction methods. Inhibition of the PCR also was observed when amplifying *Borrelia burgdorferi* DNA from blood-fed ticks (Schwartz et al. 1997).

Some PCR protocols require higher quality DNA than others. For example, amplified fragment length polymorphism PCR (AFLP-PCR) allows insects to be "fingerprinted" (see below), but requires very pure and high-quality DNA that can be cut *completely* by restriction endonucleases. When different DNA extraction methods were compared, two of the complex methods failed to produce adequate *amounts* of DNA, one simple method produced only poor *quality* DNA, but three treatments (two complex DNA methods involving phenol treatments plus a CTAB-based protocol) produced an adequate quality and quantity of DNA (Reineke et al. 1998).

The PCR is inhibited by a variety of impurities, including complex polysaccharides, heme in blood, humic substances in soil, proteases, urea in urine, phenol, and detergents (Schwartz et al. 1997, Al-Soud and Radstrom 1998). One solution to reducing the amount of impurities is to dilute them (Table 8.2). Upon dilution, however, the template DNA must remain sufficiently abundant. Other methods for eliminating inhibitors include the use of dialysis or centrifugation in cesium chloride gradients, but these methods can result in the loss of large amounts of the template DNA.

Experimental procedures sometimes can introduce inhibitors of the PCR. For example, Lee and Cooper (1995) found that PCR carried out on DNA cloned into *E. coli* failed when the bacterial colonies containing the clones were picked from plates with wooden toothpicks. The nature of the water-soluble inhibitor in the wooden toothpicks is unknown, but the toothpicks negatively affected both *Taq* and Vent DNA polymerases.

If inhibition is a serious problem, it might be reduced by embedding whole cells in low-melting-point agarose blocks, then immersing the blocks in a lysis buffer, which results in intact genomic DNA with minimal shearing damage. The agarose is then washed and cellular debris and other contaminants diffuse out during the lysis and washing steps, resulting in highly purified genomic DNA free of contaminants (Moreira 1998). The agarose-embedded

DNA can be used directly because the PCR is unaffected by the presence of high-quality low-melting-point agarose in concentrations up to 0.3%.

The use of degraded and fragmented DNA as a template for the PCR generally should be avoided because it reduces the efficiency of the PCR and limits the size of the products that can be amplified (Golenberg et al. 1996).

Typically, 0.05 to 1.0 μg of genomic DNA is needed to amplify single-copy genes with the standard PCR. Less DNA (0.5 to 2 ng) can be used to amplify multiple-copy genes such as nuclear ribosomal RNA genes, because these genes are repeated about 200 to 500 times in the eukaryotic genome. It even may be possible to conduct PCR on ancient or degraded DNA if the template is "reconstructed" (Golenberg et al. 1996).

Table 8.3 describes a standard protocol for preparing genomic DNA suitable for the PCR from a single *D. melanogaster* (Jowett 1986). Other techniques are possible (Table 8.4).

8.3.9. PCR Automation

The PCR involves repeated cycles with at least two, and generally three, temperatures (Figure 8.2). A high temperature is needed to denature (separate the two strands) the DNA template and subsequent product molecules. The lower temperature should allow annealing of the primer to the denatured ss DNA template. A third, intermediate temperature near the optima for DNA polymerase function is used for the extension (synthesis of the complementary sequence) phase. The temperature cycles can be achieved by moving samples between two or more water baths, but this is extremely tedious and it is difficult to maintain precise control over timing.

The annealing temperatures should be neither too low nor too high. An algorithm can be used to determine the optimal annealing temperature for a given pair of primers and

Table 8.3. Extracting Genomic DNA from a Single *Drosophila melanogaster*

1. In a 1.5 ml microfuge tube freeze a fly in liquid nitrogen. (Store at −70°C until needed.)
2. Thaw and add 100 μl of 10 mM Tris-HCl (pH 7.5) 60 mM NaCl, 50 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine.
3. Grind fly with a yellow pipette tip.
4. Add 100 μl of 1.25% SDS, 0.3 M Tris-HCl, 0.1 M EDTA, 5% sucrose, 0.75% freshly added DEP (diethylpyrocarbonate).
5. Mix and incubate 30 to 40 min at 60°C to ensure lysis of the nuclei.
6. Cool and add 30 μl of 8 M potassium acetate.
7. Cool for 45 min on ice.
8. Spin for 1 min in a microfuge.
9. Remove supernatant, avoiding the lipid on the surface, and add 2 volumes of ethanol.
10. Leave at room temperature.
11. Spin for 5 min and pour off the supernatant.
12. Wash the pellet with 70% EtOH.
13. Dry under vacuum.
14. Take up pellet in 25 μl of TE (10 mM Tris, 1 mM EDTA).

The method lyses the nuclei once the tissue is broken up. The SDS and protein form complexes. Diethylpyrocarbonate (DEP) is a protein denaturant and nuclease inhibitor. The protein/SDS complexes are precipitated by adding potassium, leaving the DNA in solution. The final DNA is contaminated with RNA, which can be removed by adding RNase to a concentration of 100 μg/ml.

Adapted from Jowett (1986).

Table 8.4. A Rapid Method for Extracting DNA from a Single Insect or Mite using Chelex 100 Chelating Resin

1. Add a single insect or mite to a microcentrifuge tube. Insects can be alive or frozen at −80°C.
2. Tap tube sharply to move the insect to the bottom of the tube. If the insect or mite is difficult to detect visually, add a small amount of buffer and spin in a microfuge tube to ensure the specimen is at the bottom.
3. Immerse the bottom of each microfuge in liquid nitrogen. Freeze a plastic pestle in liquid nitrogen. (Pestles are prepared in advance by melting the ends of 200-μl pipettor tips and molding them to the bottoms of microcentrifuge tubes.)
4. Macerate the frozen specimen well within the tube with the frozen pestle.
5. Add 200 μl of a 5% (w/v) Chelex solution (Bio-Rad Laboratories).
6. Vortex the solution vigorously to thaw it.
7. Remove the pestle and place tube into a temperature cycler and heat to 56°C for 15 min.
8. Centrifuge the sample (>100 g) for 15 sec in a nanofuge to allow removal of the DNA solution from the top of the tube. Avoid removing any Chelex resin from the bottom of the tube. The DNA can be used for both traditional and RAPD-PCR. The DNA is not suitable for cutting with restriction enzymes, ligation reactions, or DNA sequencing. The DNA can be stored for a few days only at −20°C.

Adapted from Edwards and Hoy (1994).

template DNA based on the GC content of the primer–template sequences (Rychlik et al. 1990). This algorithm is available in computer programs (Osborne 1992).

A variety of commercial thermal cyclers controlled by microprocessors automate the rapid and precise heating and cooling required for maximum efficiency of the PCR. Three basic categories of commercial temperature cyclers are available in which the reaction is: 1) heated and cooled by fluids, 2) heated by electric resistances and cooled by fluids, or 3) heated by electric resistances and cooled by semiconductors. Accuracy and reproducibility in temperature control should be a concern when choosing a temperature cycler. Temperature cyclers are designed for use with 0.5- or 0.2-ml microfuge tubes or with 96-well plates.

Maintaining a close fit between the walls of the block and the microfuge tube or well plate is critical in maintaining accurate temperatures. Transfer of heat can be encouraged by prefilling the wells with glycerol or mineral oil. Temperature cycles also can be achieved in an oven, which will allow nonstandard containers to be used for the PCR. Thermocyclers should be checked periodically to determine their accuracy and calibrated if needed; differences of even 1 to 2°C can be significant.

Commercially available temperature cyclers cost approximately US $3000 to $8000. They differ in the design of the cooling system, control of ramping time between temperature steps, memory capacity for program storage, sequential linking of programs, and capacity of the heating block to hold different numbers of samples for amplification. Machines that do not provide a uniform temperature across a heating block can lead to variation in outcomes from the reactions taking place in different samples. Different models or brands of temperature cyclers, although ostensibly programmed to produce the same temperature profiles, may not be equivalent and also can alter the outcomes of the PCR.

8.3.10. Specificity of the PCR

The specificity (or fidelity) of PCR based on DNA synthesis by the Klenow DNA polymerase is low. The use of *Taq* and other DNA polymerases not only simplifies the PCR but increases the specificity and overall yield. The higher temperature optimum for *Taq* (ca. 75°C) allows the use of higher temperatures for primer annealing and extension, which increases the

stringency of the reaction and minimizes the extension of primers that are mismatched with the template.

The increase in specificity with the use of *Taq* also results in an increased yield of the target fragment because competition by nontarget products for DNA polymerase and primers is reduced. In the later cycles, the amount of polymerase may no longer be sufficient to extend all the annealed primer–template complexes in a single cycle interval, which results in reduced efficiency and a "plateau." This plateau is reached after approximately 30 cycles when *Taq* is used rather than after 20 when the Klenow fragment is used.

Modifications of the standard PCR can enhance the outcome. Stringency of the annealing step can be controlled by adjusting the annealing temperature; high-temperature annealing and extension (greater than 55°C) and a balanced ratio of Mg^{2+} and dNTP concentrations give the greatest fidelity in the final product (Table 8.2). Various additives such as DMSO (2 to 5%), PEG 6000 (5 to 15% polyethylene glycol), glycerol (5 to 20%), nonionic detergents, and formamide (5%) can be incorporated into the reaction to increase specificity (Roux 1995).

Optimizing the annealing temperature and minimizing the incubation time during the annealing and extension step limits the amount of mispriming. Reducing primer and *Taq* concentration also reduces mispriming. Changing the $MgCl_2$ concentration can increase specificity by allowing a higher annealing temperature, which increases the stringency of the reaction.

Although *Taq* has no 3′ to 5′ exonuclease (proofreading) activity, its error rate is relatively low compared to that of the Klenow fragment (Mullis et al. 1986, Keohavong and Thilly 1989), because *Taq* has a 5′ to 3′ exonuclease activity during polymerization (Erlich et al. 1991). Current estimates of misincorporation rates are 10^{-5} nucleotides per cycle under optimized conditions. *Taq* appears to extend a mismatched primer–template significantly less efficiently than a correct primer–template. Misincorporated bases cannot be removed, and this can promote termination of the extending DNA chain, which will prevent propagation of the errors in subsequent PCR cycles but lowers the yield of PCR products. Because the accumulation of mutations in the PCR product is proportional to the number of DNA replications, the fewer cycles that are required to provide an adequate yield of DNA the better. Starting with adequate amounts of template (but not too much) reduces the number of PCR cycles required to produce a specific amount of product, and hence the number of misincorporations.

Sometimes variability in PCR assays is due to bad batches of *Taq* DNA polymerase, but a functional assay can be carried out to test its performance (Wada et al. 1994).

False negative results can occur for no apparent cause. These may be due to "interferences between our target DNA and the rest of the genome" (Baldrich et al. 1999). A solution can be to first digest the genomic DNA with a restriction enzyme that cuts outside the target region, followed by electrophoresis of the digested DNA, followed by recovering the restriction fragments of approximately the desired size by elution from the agarose gel. These fragments are then used as the template.

8.3.11. Detecting Primer Artifacts

Low-molecular-weight DNA artifactual products may be produced and are most obvious if the PCR is carried out with high primer concentrations, too much *Taq* in early cycles, small amounts of template DNA, and too many cycles. The artifacts may be "primer-dimers"

or other artifacts derived from the primers. Methods have been developed to eliminate primer-dimer accumulation (Brownie et al. 1997).

Primer-dimers occur when the enzyme makes a product by reading from the 3′ end of one primer across to the 5′ end of the other. As each primer serves as both primer and template, a sequence complementary to each primer is produced, which upon denaturation is a perfect template for further primer binding and extension. As the number of cycles is increased over 30, the probability of mispriming increases, as does the amount of artifact formed. The accumulation of a large amount of primer-dimer depletes primers and dNTPs from the reaction mixture and competes for enzyme with the desired target DNA.

If a PCR produces inadequate amounts of product, conducting a second amplification is a better solution to the problem than increasing the number of cycles of a single PCR. The second reaction is best done using 1 μl of the first reaction as template and a fresh reaction mixture.

8.3.12. How Many Cycles Does a PCR Need?

The answer is not too many and not too few. The optimum varies with the starting concentration of the template, the quality of the template, and the amount of inhibitory substances in the reaction, as well as all the other parameters (Table 8.2).

Too many cycles can increase the amount of nonspecific background products. Too few cycles will give a low yield that can't be detected upon electrophoresis and staining with ethidium bromide. Too much template DNA actually can inhibit the PCR. Too little template DNA can result in false negatives (Rameckers et al. 1997). The number of template molecules and cycles needed to give a good yield (about 10 ng of DNA) was estimated, making the assumption that the efficiency of the PCR actually is approximately 70% (not 100%) and the product is 200 bp in length (Rameckers et al. 1997):

Number of template DNA molecules	Theoretical number of cycles required
1	44
10	40
100	35
1,000	31
10,000	27
100,000	22

The efficiency of DNA amplification declines in the later cycles. This is called the **amplification plateau** because the product stops being produced exponentially and enters a linear or stationary phase (Kainz 2000). The plateau appears to be due to the binding of DNA polymerase to its amplification products. In general, it is better to set up multiple reactions if large amounts of DNA are needed.

8.3.13. Reducing the Evils of Contamination

It is crucial that laboratory techniques be meticulous to prevent contamination of the laboratory, supplies, and equipment with target DNA. Contamination can be an enormous problem because allele-specific PCR can generate copies of DNA from very small amounts of template (theoretically from a single molecule). Carryover of tiny quantities of PCR product can lead to **false positives** in subsequent reactions.

There is no simple and guaranteed method to prevent contamination. You must be thoughtful and careful *at all times*, employing a variety of approaches to *reduce* the possibilities of contamination. Most importantly, you must have adequate controls to *detect* contamination.

Work surfaces can be decontaminated with 0.07 M sodium hypochlorite (10% bleach) or a commercial product such as RNase solution (which eliminates both DNA and RNA) on a regular basis. UV irradiation of the workstation can be helpful, although dried DNA is less susceptible to UV irradiation than hydrated DNA (Roux 1995). Don't forget that UV irradiation of DNA polymerases and primers can damage them, reducing the efficiency of the PCR. UV light was reported to inhibit PCR amplification efficiency, even when only the water was irradiated, so routine decontamination with UV light should be used cautiously (Pao et al. 1993).

Autoclaving may not eliminate DNA contamination (Dwyer and Saksena 1992). In fact, PCR protocols published by the Cold Spring Harbor Laboratory recommend using microfuge tubes and tips without autoclaving them first to reduce the likelihood that undegraded DNA left over from previous autoclave cycles will contaminate them (Sambrook and Russell 2001).

It is *crucial* to separate physically the PCR amplification site from the location where the PCR products are evaluated by electrophoresis. Ideally, three separate sites or rooms will be available, one for DNA extraction, one for PCR amplification, and one for analysis of PCR products. Each separate room or containment unit should have a *separate* set of supplies and pipettors. Amplified DNA should never be brought into the area where DNA is being prepared for amplification or where it is being extracted. Reagents and supplies should never be taken from an area where PCR analyses are performed.

PCR reagents should be aliquotted to minimize the possibility of contamination. All reagents should be prepared, aliquotted, and stored in an area free of PCR products. Similarly, primers should be synthesized and purified in an environment free of PCR products.

To reduce contamination from barrels of pipettors, use positive-displacement pipettors with disposable tips and plungers that are completely self contained. Don't "shoot" the tips off after use; that helps to make an aerosol of the DNA. Gently pull tips off the pipettor, especially after handling PCR products. Tips that are plugged with a filter should reduce contamination of the pipettor from DNA aerosolization. Contamination also can come from electrophoresis equipment, dot-blot apparatus, razor blades, microcentrifuges, water baths, and other equipment.

Contamination risks can be reduced by changing gloves frequently (especially between DNA extraction, PCR amplification, and analysis), wearing different laboratory coats for DNA preparation, PCR amplification, and analysis, uncapping tubes carefully to reduce aerosol formation, and minimizing handling of DNA samples (Kitchin et al. 1990). Components of the PCR (mineral oil, dNTPs, primers, buffer, and enzyme) can be added to the tubes before adding the target DNA. Contamination of the tube will be reduced if each tube is capped before DNA is added to the next.

The use of positive controls can create a contamination problem. Ideally, if a **positive control** is necessary to demonstrate that your PCR is working appropriately, it should consist of template DNA that amplifies *weakly*, but consistently. Using DNA that produces strong positive responses will generate large amounts of amplified DNA, which is likely to cause contamination problems. It may be undesirable, and unnecessary, to use a positive control after the PCR has been optimized.

By contrast, multiple **negative controls** *always* should be included in the PCR experiment because they will allow you to detect contamination (**false positives**) if used consistently and in adequate numbers. Negative controls consist of all reagents but lack template DNA. A small number of contaminating template DNA molecules in the negative controls could lead to *sporadic* false positive results. Thus, it is important to carry out multiple reagent (negative) controls *each time* so that rare contaminants can be detected. How many negative controls should be used? There currently is no standard number, but a statistician might say that having *more* negative controls than experimental units is desirable.

In conclusion, although UV irradiation has been recommended as effective in inactivating contaminating DNA and has been widely used, it should not be counted on as the only method to prevent contamination (Dwyer and Saksena 1992, Frothingham et al. 1992). Furthermore, autoclaving may not eliminate previously amplified PCR products (Dwyer and Saksena 1992). Thus, meticulous attention to the entire set of procedures, including the physical separation of DNA isolation, PCR amplification, and PCR analysis, is critical in minimizing contamination or carryover problems.

8.4. Some Modifications of the PCR

Up to now, the discussion has described allele-specific PCR for which primers can be designed because sequence information is available. What can you do if you want to amplify DNA from an arthropod for which little genetic information is available? What if you want single-stranded DNA rather than double-stranded DNA as a product? What if you want to find the sequence of DNA upstream or downstream from a specific gene?

Some solutions to these, and other, problems have been achieved by modifying the types and numbers of primers used in the PCR (Table 8.5).

8.4.1. AFLP for DNA Fingerprinting

Amplified Fragment Length Polymorphism (AFLP) PCR provides a method for developing DNA fingerprints that eliminates some of the problems inherent in RAPD-PCR, AP-PCR, and DAF (described below) (Savelkoul et al. 1999). AFLP-PCR has its own limitations, namely, that it generates dominant rather than codominant markers (Mueller and Wolfenbarger 1999) and it requires absolutely clean template DNA in consistent quantities.

As will be described below, RAPD-PCR, AP-PCR, and DAF fingerprinting methods are based on amplifying random genomic DNA fragments using arbitrarily selected PCR primers, which means that "DNA fingerprints" can be generated from any DNA without prior knowledge of the DNA sequence. These PCRs are performed at low annealing temperatures to allow the primers to anneal to the template at multiple loci, which makes them very sensitive to reaction conditions, DNA quality, and PCR temperature profiles.

AFLP-PCR eliminates most of these problems because it is based on detecting restriction fragments by PCR amplification; AFLP-PCR can be used on DNAs of any origin and complexity, without requiring prior knowledge of sequence, and using a limited set of generic primers (Vos et al. 1995, Savelkoul et al. 1999).

There are three steps in AFLP-PCR: 1) template DNA is digested by restriction enzymes, and oligo "adapters" are ligated to the digested DNA; 2) sets of restriction fragments are selectively amplified using the adapter and restriction site sequences as target sites for primer annealing; 3) the amplified fragments are analyzed by electrophoresis.

Table 8.5. Some Modifications of the PCR Use Different Types of Primers

PCR type Primer number, nt length	Potential uses
Standard allele-specific	
Paired, 15 to 30 nt each	Amplify DNA for which sequence information is available for the target DNA.
AFLP	
Generic primers based on restriction site sequences and "adapter" sequences	DNA is digested by restriction enzymes and oligo "adapters" are ligated to digested DNA and amplified using generic primers that use the restriction site sequences and adaptor sequences as target sites for primer annealing. AFLP-PCR yields multiple fragments.
Anchored	
One known primer, second is made	Amplify DNA when only one prime sequence is known. Synthesis of cDNA with the known primer is carried out using mRNA; a polyG tail is added to the cDNA. The second primer is made by synthesizing a primer with a polyC sequence, which allows amplification of a second DNA strand that is complementary to the cDNA.
Arbitrary	
Single, 18- to 30-nt arbitrary sequence	Amplify regions of DNA internal to regions to which arbitrary primers (such as M13 sequencing primer, M13 reverse sequencing primer, or T3 sequencing primer) anneal on opposite strands. One or more DNA fragments will be produced, and these can be used to generate genome maps or discriminate between individuals, populations, or species.
Asymmetric	
Paired, 10 to 30 nt in a 1:50 to 1:100 ratio	Amplify ss -stranded DNA for sequencing.
Degenerate	
Multiple types, 15 to 30 nt	Amplify DNA that is related to genes for which the sequence or part of the sequence is known in a related species, or for members of a gene family. The degeneracy of the DNA code, and codon bias, for amino acids determines how many primer types are needed in the reaction.
Inverse	
Paired, 15 to 30 nt inverse orientation	Amplify regions of DNA of unknown sequence that flank known sequences; used for identifying upstream/downstream sequences. Primers are oriented so DNA synthesis occurs away from the known "core" DNA.
Multiplex	
Multiple primers	More than one pair of primers amplify several DNA targets simultaneously. Careful optimization of PCR conditions is required to produce consistent results.
PCR-RFLP	
Paired, 18 to 30 nt	Nuclear DNA is amplified by the standard PCR, then the product is cut with restriction enzymes. Banding patterns are visualized on a gel after staining with ethidium bromide.
Quantitation of mRNA	
Paired, 15–30 nt	Several methods: 1) Two different cDNAs are amplified, and the absolute level of one is calculated if the other is known. 2) The sample is spiked with a known amount of control DNA, and target and control DNA are amplified and compared to estimate amount of target DNA.
RAPD-PCR	
Single, 10 nt random sequence	Random amplified polymorphic DNA PCR. Amplify regions of DNA that are flanked by the random primer sequences. Multiple DNA fragments may be produced and used as markers for genome mapping or identifying individuals, populations, or species.
RNA amplification	
18 to 22 nt	mRNA is reverse transcribed and the cDNA is amplified by PCR.

The number of fragments produced in a single AFLP-PCR can be determined by selecting specific primer sets. Annealing conditions can be stringent in AFLP. AFLP-PCR allows analysis of closely related populations and species. For example, European populations of an introduced pest, the grape phylloxera, were compared by AFLP-PCR with North American populations (Forneck et al. 2000). Two distinct populations, northern and southern, were found in Europe, and AFLP-PCR patterns suggest that two different introductions occurred, one from the northeastern United States and the other from the south-central United States.

A simplified version of AFLP-PCR was developed to discriminate between European and African honey bees (Suazo and Hall 1999). The protocol involved digesting DNA and ligating the adapters in one reaction rather than two; one restriction enzyme was used rather than two; and amplification was accomplished in one reaction rather than two. Finally, the PCR products were electrophoresed in agarose-Synergel instead of polyacrylamide and visualized by ethidium bromide staining rather than autoradiography of labeled primers. These modifications in AFLP-PCR may reduce the amount of polymorphism detected.

8.4.2. Anchored PCR

In situations in which only one sequence is known that is suitable for a primer (rather than two), **anchored PCR** can be used. The procedure involves synthesis of cDNA with the known primer from mRNA (Collasius et al. 1991). A polyG tail is added to the cDNA. The second primer is developed by synthesizing a primer with a polyC sequence, which allows amplification of a second DNA strand that is complementary to the cDNA. Subsequent cycles yield amplified DNA from both strands.

8.4.3. Arbitrary Primers

Ecologists, evolutionary biologists, and geneticists often wish to develop genetic markers for insects for which little genetic information is available. **Arbitrarily Primed PCR** (AP-PCR) can produce a characteristic "fingerprint" pattern for any genome, which could be useful for developing markers for breeding programs, genetic mapping, population genetics, or epidemiology (Welsh and McClelland 1990, Welsh et al. 1992, McClelland and Welsh 1994).

AP-PCR involves two cycles of low-stringency amplification, followed by cycles conducted at higher stringency, using a single primer of arbitrary sequence. The term **stringency** refers to PCR conditions such as the annealing temperature. If a high annealing temperature is employed, then the primers will only anneal to the template DNA if a high proportion of the sequences match. Lower annealing temperatures allow some mismatches. Full-length primers (20 to 34 nt long) that have been used include the Universal M13 sequencing primer, the M13 reverse sequencing primer, and the T3 sequencing primer.

How does AP-PCR work? At lower temperatures, an arbitrary primer can anneal to many sequences with some mismatches. By chance, some primers will be able to anneal to the target DNA within a few hundred bases of each other and on opposite strands. Sequences between these positions then will be amplified. The extent to which sequences amplify depends on the efficiency of priming and the efficiency of extension. During early cycles, those sequences that prime most efficiently will predominate. During later cycles, those that amplify most efficiently will predominate.

Between three and 20 DNA products typically are produced in AP-PCR, which allows differentiation between closely related strains of some species (Welsh et al. 1990).

AP-PCR also has been used to amplify RNA in order to detect and clone mRNAs that are differentially expressed in different cells (McClelland et al. 1995). Clones of the aphid *Ceratovacuna nekoashi* from a single gall were shown by AP-PCR to be genetically identical, whereas aphids from different galls on the same twig were successfully differentiated, indicating that members of a gall constitute a clonal population, a gall is founded by a single female, and intergall migration is absent or rare (Fukatsu and Ishikawa 1994).

A modification of AP-PCR was developed and called DALP, or Direct Amplification of Length Polymorphisms (Desmarais et al. 1998). DALP uses the M13 sequencing forward primer as a core sequence for the forward primer and the M13 reverse primer. These primers produce specific multibanded patterns that show interindividual length variations. Each band then can be sequenced with the universal sequencing primers.

8.4.4. Asymmetric PCR

Single-stranded DNA can be produced by **asymmetric PCR**. By providing an excess of primer for one of the two strands, typically in ratios of 50:1 to 100:1, amplification results in product that is primarily single-stranded. Early in the reaction, both strands are produced, but as the low-concentration primer is depleted, the strand primed by the abundant primer accumulates arithmetically. Such ss DNA is particularly useful for sequencing (see Chapter 7).

8.4.5. Degenerate Primers

If only a limited portion of a protein sequence is known for a target gene, **degenerate primers** may allow detection of new or uncharacterized sequences in a related family of genes, or may amplify members of a gene family. Degenerate primers are a mixture of oligonucleotides varying in base sequence, but with the same number of bases.

Designing degenerate primers for the PCR requires several considerations. You will recall that the genetic code is degenerate (with more than one codon for most amino acids). Methionine and tryptophan are encoded by a single codon, but the other amino acids are encoded by two to six different codons. When designing degenerate primers, it is useful to chose a segment of the protein in which the amino acids have minimal degeneracy. The lower the degeneracy in the primers, the higher the specificity of the PCR. The degeneracy of the primer may be restricted further by considering which codons are most often used in a particular species (codon bias), if it is known. Furthermore, degeneracy may be reduced if primers containing fewer (15 to 20) nucleotides are used. Because a single mismatch, especially at the 3′ end of the primer, may prevent *Taq* from extending, degeneracy at the 3′ end should be avoided. Empirical testing of primers may be necessary and modifications made to ensure that the desired product is synthesized.

8.4.6. Hot-Start PCR

A hot-start PCR protocol can optimize the yield of the desired product while limiting the likelihood of nonspecific amplification. Hot-start PCR is achieved by leaving an essential component out of the reaction mixture until the mixture has been heated to a temperature that inhibits nonspecific priming and extension. Typically, all PCR components are added and held at high temperature before the DNA polymerase is added.

A modification of this method involves using wax to provide a physical barrier between the components of the reaction. The primers, Mg^{2+}, dNTPs and buffer can be mixed

at room temperature in the bottom of the reaction tube and then covered with melted wax that melts at low temperature (53 to 55°C). The remaining components are then added on top of the wax barrier. During the first cycle of the PCR the wax barrier melts during the denaturation step, allowing the components to combine. The melted wax floats to the top of the mixture where it acts as a barrier to evaporation. Hot-start *Taq* DNA polymerase is now available in which the enzyme is activated only after the reaction reaches 94°C or higher, allowing all components to be mixed at room temperature.

Hot-start PCR especially is useful when nonspecific amplification is a problem because there is too little template DNA, the template DNA is complex, or several pairs of primers are used (multiplex PCR) (Sambrook and Russell 2001).

8.4.7. Inverse PCR

An unknown sequence that flanks a "core" region with a known sequence can be amplified by **inverse PCR** (Ochman et al. 1990, Sambrook and Russell 2001). Inverse PCR involves digesting the template DNA with a restriction endonuclease that cuts outside the region of known sequence to produce a fragment about 3 to 4 kb in length. Southern blot analyses may be necessary to identify restriction enzymes that produce fragments of suitable size for circularization and amplification. If an enzyme is used that cleaves within the core region, either the upstream or downstream segment of DNA will be amplified.

Once the DNA has been digested, the ends of the fragment are ligated to form a circular molecule. Ligation is performed with T4 DNA ligase in a dilute DNA concentration to favor formation of monomeric circles. Amplifying the flanking DNA outside the core region is carried out using primers oriented in the direction opposite to the usual one (Silver 1991). Primers for inverse PCR thus are homologous to the ends of the core region so that DNA synthesis proceeds across the *uncharacterized* region of the circle rather than across the characterized core region (Figure 8.3).

8.4.8. Long PCR or High-Fidelity PCR

In theory, standard PCR using *Taq* DNA polymerase can amplify up to ~10 kb of DNA sequence. In practice, amplification of such long sequences is difficult, and most amplifications are limited to 1 to 2 kb. However, large (up to 40 kb) DNA fragments to be amplified with high fidelity and yield if two DNA polymerases are used (Barnes 1994, Cheng et al. 1994a,b).

The effectiveness of Long PCR is due to the combined action of two DNA polymerases. A polymerase such as *Pfu* has a 3′ exonuclease activity (which *Taq* does not have), which means that *Pfu* can remove any accidental mismatches introduced into the growing DNA molecule. The theory is that one of the deterrents to truly long PCR products may be due to the incorporation by *Taq* of mismatched or damaged nucleotides every few kilobases, thus eliminating those DNA molecules from further amplification. The incorporation of errors occurs because *Taq* lacks a proofreading function. Adding a small amount of a proofreading polymerase to an excess of *Taq* provides *Taq* with a proofreading "helper." Primer mismatches are corrected, and *Taq* is able to extend primers for longer distances. Several combinations of DNA polymerases have been used in Long PCR; Cheng et al. (1994a) found that rTth (from *Thermus thermophilus*) and Vent (from *Thermococcus litoralis*) polymerases were the most reliable combination under their test conditions.

Several other modifications in the Long PCR protocol enhance the likelihood of obtaining longer products. Long PCR typically uses long primers (at least 30 nt), works best at a higher

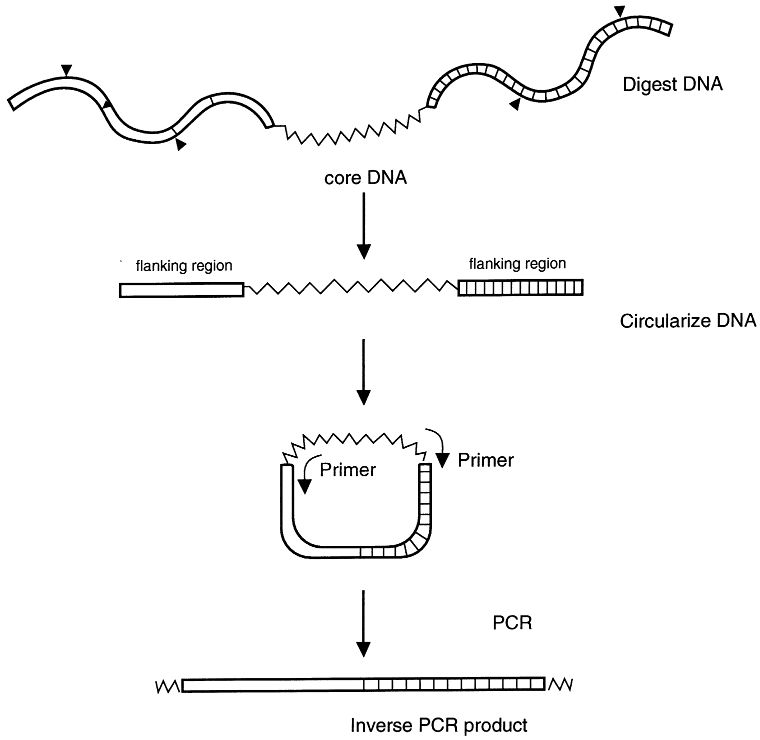


Figure 8.3. Inverse PCR allows amplification of DNA flanking the "core" DNA, for which sequence information is available. Step 1 involves digesting the template DNA with an appropriate restriction enzyme to produce fragments approximately 2–4 kb long, with the core DNA in the middle. Step 2 involves circularizing the DNA by ligation. Primers, dNTPs, and DNA polymerase are added and the PCR is carried out. Primers are oriented so that synthesis of DNA occurs away from the core DNA into the flanking regions. The PCR product consists of the two flanking regions.

pH (8.8 to 9), and uses a high rate of change in temperatures (rapid cycling) and a longer synthesis interval.

Why not simply use a single DNA polymerase with 3′-exonuclease activity to edit out the mismatches in extension? Barnes (1994) suggested that the enzymes with 3′ exonuclease activity can degrade PCR primers, especially during the long synthesis times. Thus, only small amounts of polymerase with 3′-exonuclease activity should be used. In addition, it is especially important that the template strands be completely denatured at high temperatures to prevent renaturation before primers can anneal and be extended (Cheng 1995).

Primer design for Long PCR, as usual, should avoid the potential for secondary structure and dimer formation (Cheng et al. 1994a). Primers 21 to 34 nt long that have melting temperatures near 65 to 70°C permit the use of higher annealing temperatures to enhance reaction specificity. Thermal cycling profiles in Long PCR typically use a hot start at 78 to 80°C, initial denaturation at 94°C for 1 min, 25 to 40 cycles of denaturation at 94°C for 15 sec, and annealing and extension steps at 60 to 68°C for 30 to 60 kb of target DNA. Typically, extension times are increased for each subsequent cycle to facilitate production of long product molecules.

Other factors that influence the success of the Long PCR include the integrity of the target DNA, which means that DNA extraction methods must be carefully considered. Longer targets can be amplified best from DNA with little shearing damage (Cheng and Kolmodin 1998). Several DNA extraction methods produce large DNA fragments from insects (Ebert 1996), although shearing of DNA may be difficult to avoid when extracting DNA from adult insects because their exoskeleton can damage the DNA during grinding. One solution is to extract DNA from embryos (Rabinow et al. 1993).

The Long PCR protocol has been used for another application—amplifying microbial DNA when mixed with arthropod DNA (Jeyaprakash and Hoy 2000, Hoy et al. 2001). When insect and microbial DNA are mixed, efficient amplification of the microbial DNA appears to be inhibited, for unknown reasons. For example, the ability to detect the endosymbiont *Wolbachia* within the bodies of various arthropods was greatly enhanced when the Long PCR, rather than a standard PCR, protocol was used; *Wolbachia* were found in 76% of the 63 arthropods examined in 13 orders. The Long PCR thus can be used to increase sensitivity or fidelity of the PCR even when shorter DNA targets are amplified.

The Long PCR protocol is approximately five to seven orders of magnitude more sensitive in amplifying *Wolbachia* DNA than the standard PCR (Jeyaprakash and Hoy 2000). When standard and Long PCR protocols were compared using known amounts of *Wolbachia* template DNA mixed with known amounts of insect DNA, the Long PCR could amplify as few as 100 copies of *Wolbachia* DNA consistently. By contrast, standard PCR was only able to detect *Wolbachia* DNA reliably when at least 100 million copies of plasmid DNA were present (Figure 8.4).

Long PCR should have many applications whenever long DNA fragments are useful; for example, Long PCR has been used to develop rapid restriction maps of DNA fragments 8 to 18 kb in length (Her and Weinshilboum 1995). Long PCR can be used to clone large genes or be a labor-saving alternative for studying larger genome segments such as entire mitochondria that are 16 to 20 kb in length (Nelson et al. 1996).

8.4.9. Multiplex PCR

More than one pair of primers can be used to amplify multiple PCR products (Sambrook and Russell 2001). In **multiplex PCR** (Figure 8.5), the goal is to amplify several segments of target DNA simultaneously, which should reduce time, minimize costs, and increase efficiency. In reality, however, the yield of each product usually is reduced in proportion to the number of primer pairs included in the reaction. Up to eight primer pairs have been used simultaneously before the yield of each product is too low to be visualized by staining with ethidium bromide on an agarose gel.

Despite the potential benefits, multiplex PCR is difficult to develop because all the primers must have approximately the same melting temperature, the primers should not interact with one another, and the amplified products should be of approximately the same size, but still be distinguished from each other by gel electrophoresis.

Steps to develop multiplex PCR include the following: 1) Determine that all target DNA can be amplified efficiently using the same PCR temperature profile. 2) Titrate the amount of each primer pair to achieve maximum amplification in separate reactions using the same program and reaction conditions. 3) Balance the amount of each primer pair to achieve acceptable levels of amplification of all targets in the multiplex reaction. One solution to problems with step 3 usually involves increasing progressively the concentration of the nonworking primer pairs while reducing the concentration of the effective primer pairs.

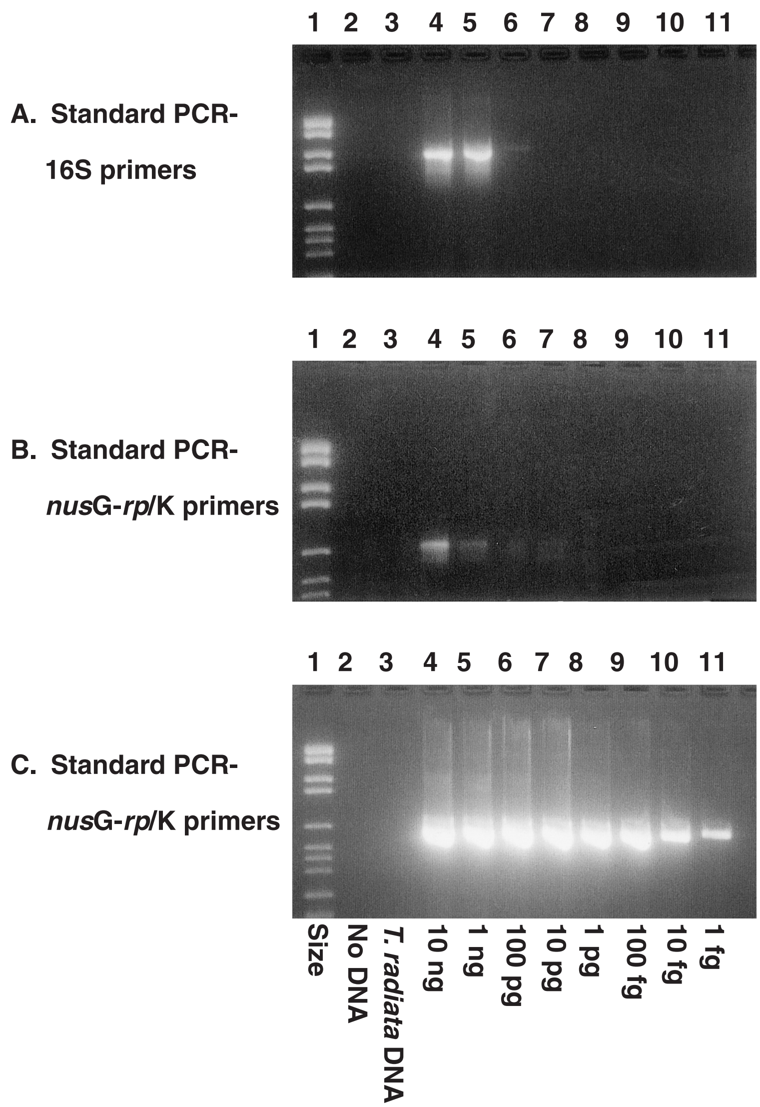


Figure 8.4. Long (High Fidelity) PCR is more sensitive by approximately six orders of magnitude than standard, allele-specific PCR when microbial DNA is mixed with insect DNA (Hoy et al. 2001). The same template DNA (a plasmid containing both the *nus*G-*rpl*K and 16S sequences) from the greening bacterium was serially diluted from 10 nanogram (ng) to 1 femtogram (fg) and added to 10 ng of parasitoid, *Tamarixia radiata*, DNA. A) 16S primers were used in a standard, allele-specific PCR protocol. Detectable products are in lanes 4 and 5 only. B) *nus*G-*rpl*K primers were used in a standard, allele-specific PCR protocol. Weak products are seen in lanes 4 and 5. C) *nus*G-*rpl*K primers were used with a Long PCR protocol. Strong products were produced in lanes 4 through 11. Lane 1, DNA size marker; lane 2, no template-DNA control; lane 3, 10 ng *T. radiata* DNA only control; lanes 3–11 contain *T. radiata* DNA + 10 ng, 1 ng. 100 pg, 10 pg, 1 pg, 100 fg, 10 fg, 1 fg of the plasmid DNA, respectively.

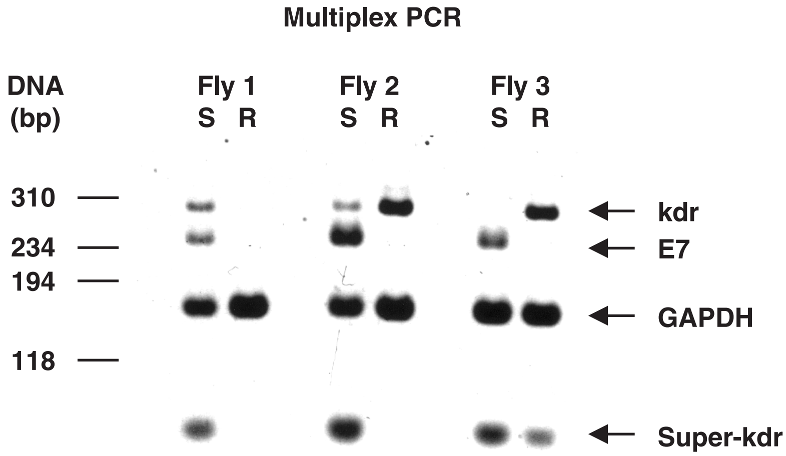


Figure 8.5. Multiplex PCR allows more than one gene to be sampled in a single reaction. In this example, three pesticide resistance genes (*kdr*, E7, and *Super-kdr*) were monitored in three hornflies, *Haematobia irritans*. Fly 1 has only susceptible alleles for *kdr*, *Super-kdr*, and E7. Fly 2 has one copy of the susceptible and one of the resistant alleles of *kdr*, but has susceptible alleles only for the E7 and *Super-kdr* genes. Fly 3 has only *kdr* alleles (is homozygous resistant) and susceptible alleles of E7 and is heterozygous for *Super-kdr*. Amplification of the GABDH gene provides a control to demonstrate that the PCR is working correctly. (Photo kindly provided by Felix Guerrero.)

Other recommendations for optimizing multiplex PCR can be found in Sambrook and Russell (2001).

8.4.10. Nested PCR

Nested PCR involves a two-step procedure in which one pair of primers is used to amplify a fragment. Subsequently, a second pair of primers is used to amplify a smaller fragment from an aliquot of the first PCR. Nested PCR is designed to be both sensitive and specific.

Nested PCR of a 16S rRNA gene from the causative agent of granulocytic ehrlichiae (*Ehrlichia chaffeensis*), a disease of humans, was found to be so sensitive that as few as two copies of the 16S gene could be detected when a spiking experiment was conducted (Massung et al. 1998). Spiking experiments involved using known quantities of a plasmid containing the 16S rRNA gene added into background human genomic DNA. The use of serial dilutions to determine how repeatable and reliable a PCR assay is should be done whenever it is important to resolve how often false negatives are likely to occur in an experiment.

8.4.11. PCR-RFLP

PCR-RFLP eliminates some of the disadvantages to traditional restriction fragment length polymorphism (RFLP) analysis for analyzing population variation using DNA isolated from individual insects (Karl and Avise 1993). If no primers are available from the literature, a genomic DNA library is constructed and clones are isolated. Clones with inserts of 500 to 2000 bp are chosen, and sequences of the first 100 to 150 nt from both ends are obtained so

that PCR primers can be derived. Nuclear DNA is amplified by the PCR using these primers and digested with appropriate restriction enzymes. The cut DNA is visualized after electrophoresis by staining with ethidium bromide. The advantage to PCR-RFLP is that DNA extracted from a single individual is sufficient, after amplification, to provide electrophoretic bands that can be visualized without having to be hybridized with radiolabeled probes.

8.4.12. Quantitative PCR

Methods have been developed to quantitate the amount of DNA or RNA present in a sample, but these remain difficult and may have high rates of error (Arnheim and Erlich 1992, Siebert and Larrick 1992, Foley et al. 1993, Sambrook and Russell 2001).

Quantitative PCR requires some form of standard with which the target sequence concentration is compared. For example, estimation of the number of amplified esterase genes in insecticide-resistant mosquitoes used a nonamplified esterase gene as an internal control (Weill et al. 2000). Quantification of infection of fleas with the plague bacterium *Yersinia pestis* was based on standard, curve-based, competitive PCR (Hinnebusch et al. 1998). This quantitative PCR method was found to be equally accurate and precise as a colony count reference method when evaluated using mock samples and laboratory-infected fleas.

Quantitative PCR is more difficult than other types of PCR due to the nature of the PCR. Because PCR is an exponential (or nearly so) process, small differences in efficiency at each cycle, especially the earliest cycles, can lead to large differences in yield. Anything that affects exponential amplification can disrupt quantitation. Thus, different amounts of inhibitors in samples containing the same amount of template DNA could result in different amounts of product, as could small differences in efficiency between the primer pairs used to amplify the standard (control) and target sequences.

Quantification of amplified products can be achieved by gel electrophoresis or by fluorescently labeled primers quantified with an automated DNA sequencer, fluorometry, analysis of gel images stained with ethidium bromide or other intercalating dyes, or measurement of radioactivity incorporated during amplification (Sambrook and Russell 2001). Another method for quantitating PCR products is through real-time PCR or TaqMan PCR, described below.

8.4.13. Random Primers

A method similar to AP-PCR was developed when Williams et al. (1990) demonstrated that genomic DNA from diverse organisms could be amplified using a single short (9 or 10 nt long) primer composed of "random" oligonucleotides (Figure 8.6).

The "random primers" can be designed without the experimenter having any genetic information for the organism being tested. The only constraints are that the primers should have 50 to 80% G+C content and no palindromic sequences. Different random primers used with the same genomic DNA produce different numbers and sizes of PCR products (Ellsworth et al. 1993, Kernodle et al. 1993, Meunier and Grimont 1993, MacPherson et al. 1993, Williams et al. 1993).

The amplified DNA can be detected as bands in ethidium bromide-stained agarose gels (Figure 8.6). This modified PCR method was called **RAPD-PCR** because it produced Random Amplified Polymorphic DNA bands.

RAPD-PCR has been used to develop genetic maps and to identify molecular markers in populations or species, as well as determine paternity in dragonflies (Hadrys et al. 1992,

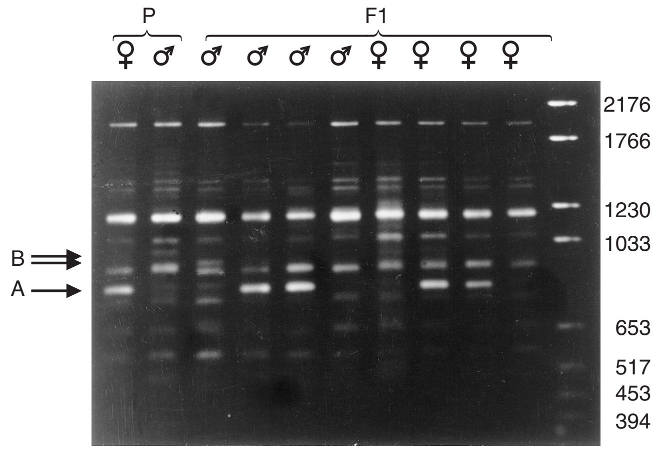


Figure 8.6. Photograph of a gel showing mode of inheritance of RAPD-PCR DNA fragments from the parasitic wasp *Trioxys pallidus*. A single 10-mer primer anneals to different regions of the genome, and if two primers anneal in the opposite orientation, amplification of several DNA sequences occurs. The size of the bands is indicated by the size marker lane on the right. Arrow A indicates that the band in the mother is inherited in about half of her haploid male and female progeny, which is consistent with the hypothesis that the mother was heterozygous for this band. Arrow B identifies faint bands that are not inherited in a Mendelian fashion and are not used in analyses (Edwards and Hoy 1994).

1993, Tingey and del Tufo 1993, Schierwater 1995). RAPD-PCR makes it possible to identify hundreds of new markers in a short time, which allows genetic maps to be developed rapidly. RAPD-PCR is particularly valuable for genome mapping in those species for which other genetic markers are lacking or rare (Laurent et al. 1998).

Genomic DNA sequences differing by only a single base may not be amplified in the RAPD protocol, or may result in a complete change in the number and size of the amplified DNA segments. Thus, RAPD-PCR may detect small differences in the genomes of individual insects or mites, different populations, or species. RAPD fingerprinting can be carried out on very small insects, such as single aphid embryos, while preserving the mother for morphometric or karyotyping analyses (Chan et al. 1999) and is especially useful for discriminating between tiny parasitoid species or biotypes (Edwards and Hoy 1993, 1994, Vanlerberghe-Masutti 1994). Differences in RAPD-PCR patterns are correlated with the evolution of different taxa, allowing limited estimates of evolutionary divergence (Espinasa and Borowsky 1998).

RAPD-PCR products can be cloned (Comes et al. 1997) and sequenced so that “allele-specific” primers can be developed for future PCR analyses. Sequence-Characterized Amplified Region (SCAR) primers will produce allele-specific PCR products. Agusti et al. (2000) used a SCAR primer pair to amplify single bands of 310 bp to detect the whitefly *Trialeurodes vaporariorum* in the gut of the predator *Dicyphus tamaninii*.

Two or more primers have been employed simultaneously to generate reproducible RAPD fragments that are different from those obtained with each single primer (Micheli et al. 1993, Sall et al. 2000).

RAPD-PCR has been criticized for its lack of reproducibility (Ayliffe et al. 1994, Lamboy 1994, Micheli et al. 1994, Hallden et al. 1996, Jones et al. 1997, Khandka et al. 1997,

McEwan et al. 1998, Perez et al. 1998). Different RAPD banding patterns can be obtained if different DNA extraction methods are used, probably due to the presence of different kinds or amounts of contaminants or different amounts of template DNA (Micheli et al. 1994). Different DNA polymerases also may amplify different RAPD products (Schierwater and Ender 1993).

RAPD-PCR is sensitive to both DNA template concentration and quality, so bands may vary in intensity or even disappear if template concentration is not controlled or DNA is sheared (Khandka et al. 1997). Reproducibility also can be poor if different PCR machines or pipettors are used, resulting in different temperature cycling conditions or different concentrations of the PCR mixture (He et al. 1994a,b, Schweder et al. 1995). Occasionally, heteroduplex molecules formed between allelic sequences can cause artifactual RAPD bands (Ayliffe et al. 1994). Thus, it is critical that researchers use primers only if they produce bright, consistent banding patterns in the particular thermocycler used. Likewise, researchers should obtain a reference profile for their own work rather than comparing their results to the those generated by another (He et al. 1994a).

Another criticism of RAPD-PCR is that all bands are inherited as dominant alleles. This means that heterozygotes cannot be identified unless progeny testing is conducted, although this is not an issue when RAPD-PCR is conducted on haploid males of arrhenotokous species (Edwards and Hoy 1993). Another problem is that comigration of similar sized bands with different sequences can occur, but may not be detected, unless the bands are cloned and sequenced.

Use of RAPD markers to calculate genetic similarity coefficients can result in false positives and false negatives if RAPD artifacts are present (Lamboy 1994). As a result, Nei and Li's coefficient is recommended for computing genetic similarities with RAPD data, particularly if PCR artifacts are present.

Some of the negative aspects of RAPD-PCR can be eliminated by a method called Sequencing With Arbitrary Primer Pairs (SWAPP) (Burt et al. 1994). In SWAPP, amplified random bands are purified from the gel, reamplified with the same two primers used in the initial amplification, and repurified. One of the primers then is added back and annealed to the product and sequenced. SWAPP allows polymorphisms in populations to be characterized at the nucleotide level, eliminates non-Mendelian inheritance, and allows bands to be produced reliably. The technique requires only small amounts of low-quality DNA and no prior genetic information on the organism.

8.4.14. Real-Time PCR

Real-time PCR can quantify gene expression and confirm differential expression of genes. Real-time PCR uses commercially available fluorescence-detecting thermocyclers to amplify specific nucleic acid sequences and measure their concentration simultaneously (Sambrook and Russell 2001). Target sequences are amplified and quantified in the same PCR machine.

Internal standards are not required in order to quantify the amount of DNA or RNA present in real-time PCR. The ability to quantify the amplified DNA during the exponential phase of the PCR, when the reaction components are not limited, results in improved precision in quantification of target sequences. Real-time PCR can measure the initial concentration of target DNA over a range of five or six orders of magnitude. At present, the limit of detection when fluorescent dyes are used is approximately 10 to 100 copies of template DNA in the

starting reaction (Sambrook and Russell 2001). The TaqMan method of real-time PCR is described below.

Significant advantages of real-time PCR include its ability to measure DNA concentrations over a large range, its high sensitivity, its ability to process many samples simultaneously, and its ability to provide immediate information. A disadvantage of real-time PCR is that it is expensive. At present, the machines are expensive, and the maintenance and operational costs are high. Currently, real-time PCR is used primarily in large commercial laboratories that process a very large number of samples of a similar type.

8.4.15. Reverse Transcription PCR

Messenger RNA can be reverse transcribed, and the resultant cDNA then can be amplified using *Taq* DNA polymerase. Reverse transcriptase-PCR (RT-PCR) allows detection of gene expression in small numbers of specific cells or tissues. Reactions have been carried out with RNA isolated from as few as 10 to 1000 cells.

The process involves: 1) isolation of mRNA, 2) reverse transcription of mRNA into cDNA, and 3) amplification of cDNA by DNA polymerase. Primers for the amplification should be 18 to 22 nt long and should occur in separate exons to inhibit amplification of any contaminating genomic DNA in the RNA preparation. Multiple reverse transcriptases can be used in RT-PCR to increase sensitivity and product yield (Nevett and Louwrier 2000).

RT-PCR has been used to monitor for the presence of rabbit hemorrhagic disease virus in fly species in Australia (Asgari et al. 1998). This calcivirus causes a lethal disease in European rabbits, but little was known about how it spreads in the field. RT-PCR provided a sensitive and reliable method for detecting the virus in flies and flyspots (feces), which allows it to be used to study the epizootiology and vector biology of the virus.

8.4.16. TaqMan PCR

TaqMan PCR is a type of real-time PCR. TaqMan PCR uses a nucleic acid probe complementary to an internal segment of the target DNA. The probe is labeled with two fluorescent moieties. The emission spectrum of one overlaps the excitation spectrum of the other, resulting in “quenching” of the first fluorophore by the second. The probe is present during the PCR and if product is made, the probe is degraded via the 5′-nuclease activity of *Taq* polymerase that is specific for DNA hybridized to template (= TaqMan activity). The degradation of the probe allows the two fluorophores to separate, which reduces quenching and increases intensity of the emitted light. Because this assay involves fluorescence measurements that can be performed without opening the PCR tube, the risk of contamination is greatly reduced. Furthermore, no electrophoresis is required, so labor and materials costs are reduced (Kalinina et al. 1997, Sambrook and Russell 2001).

8.5. Some Research Applications

The PCR can be applied to a diverse array of both basic and applied problems (Table 8.6). Protocols for the different methods are available in books (Erlich 1989, Erlich et al. 1989,

Table 8.6. Some Entomological Problems and Potential PCR Protocols

Problem	PCR technique(s)
Amplify ancient DNA	Standard allele-specific PCR
Amplify mRNA	RNA PCR
Chromosome walking	Inverse PCR; long PCR
Cloning a gene	Blunt end cloning; sticky-ended cloning; anchored PCR; PCR with degenerate primers; long PCR
Constructing a genetic map	AP-PCR; RAPD; inverse PCR
Constructing a phylogeny	Standard PCR with primers having polylinkers for cloning/sequencing; asymmetric PCR and sequencing; PCR-RFLP; multiplex PCR
Detecting gene expression	RNA PCR; TaqMan PCR
Detecting mutations	Standard PCR; RAPD; AP-PCR; PCR-RFLP
Detecting pathogens in arthropod vectors	Standard PCR; long PCR
Detecting transgenic arthropods	Standard PCR
Engineering DNA	
Introduce restriction sites into DNA fragments	Attach sequences to $5'$ end of primers and conduct standard PCR probes, or isolating DNA strands on a column
Label DNA with ^{32}P or biotin for sequencing	
Assemble overlapping DNA segments to make synthetic DNA	Alter primer sequence when synthesizing, then standard PCR
Introduce substitutions, deletions, or insertions in product DNA	
Evolutionary analyses	Standard PCR; RAPD; AP-PCR; DNA sequencing; PCR-RFLP
Identify species	Standard PCR; RAPD; AP-PCR; PCR-RFLP
Identify strains, races, or biotypes	Standard PCR; RAPD; AP-PCR; PCR-RFLP
Identifying upstream/downstream sequences	Inverse PCR; SSP-PCR
Monitoring dispersal of individuals	Standard PCR; RAPD; AP-PCR
Sequencing a gene	Asymmetric PCR to produce ss DNA; dideoxynucleotide chain-termination sequencing method with *Taq* polymerase; cycle sequencing; direct sequencing

Innis et al. 1990, Ausubel et al. 1991, McPherson et al. 1991, Sambrook and Russell 2001) or individual journal papers.

The following examples provide evidence of the versatility of the PCR, but are only an abbreviated introduction to the diversity of applications to which this tool can be applied. Modifications of the PCR continue to be made to resolve diagnostic, ecological, evolutionary, genetic, and developmental biology questions.

8.5.1. Amplifying Ancient DNA

The film *Jurassic Park* implied it was possible to amplify dinosaur DNA from insects preserved in amber; this captured the imagination of the public and created a climate in which the PCR was perceived to be an unusually powerful key to analyzing the past. Subsequently, the PCR was used to amplify DNA fragments from a number of insects preserved in ancient amber. Unfortunately, these results have been controversial (Box 8.1), as have been the results of amplifying dinosaur DNA (Austin et al. 1997a,b, Rollo 1998, Hofreiter et al. 2001).

Box 8.1. Amplifying ancient DNA from insects in amber: Controversial results?

Why the controversy? Is amber a special form of preservative that allows DNA to persist for unusually long periods of time (millions of years)? Amber entombs insect specimens completely, after which they completely dehydrate so the tissue is effectively mummified. The terpenoids, major constituents of amber, could inhibit microbial decay (Austin et al. 1997a). Certainly, preservation of amber-embedded insects seems to be exceptional and insect tissues in amber appear comparable in quality to the tissues of the frozen wooly mammoth (which is "only" 50,000 years old). But is the DNA in these tissues preserved and can it be amplified by the PCR?

Claims have been made that DNA has been extracted from a variety of insects in amber, including a fossil termite, *Mastotermes electrodominicus*, estimated to be 25 to 30 million years old (DeSalle et al. 1992); a 120- to 130-million year old conifer-feeding weevil (Coleoptera: Nemonychidae) (Cano et al. 1993a); and a 25- to 40-million year old bee (Cano et al. 1993b). These are extraordinary ages for DNA!

The DNA sequences obtained from all amber-preserved insects meet several, but not all, criteria of authenticity; the fossil DNA sequences "make phylogenetic sense" and DNA has been isolated from more than one specimen in several cases (although the weevil DNA was derived from a single specimen) (Austin et al. 1997a).

Yet extraction and amplification of fossil DNA sequences from amber-preserved insects has yet to be reproduced in independent laboratories, despite multiple attempts to do so. This has cast doubt on the authenticity of the reports (Austin et al. 1997a,b, Sykes 1997, Walden and Robertson 1997, Gutierrez and Marin 1998, Hofreiter et al. 2001).

One of the most controversial claims involved the isolation of a "living" bacterium from the abdomen of an amber-entombed bee. Bacterial DNA from a 25-million-year-old bee was obtained and sequenced and a bacterial spore was reported to be revived, cultured, and identified (Cano and Borucki 1995). The classification of the bacterium is controversial (Beckenbach 1995, Priest 1995) because the bacterium could have come from a currently undescribed species of the *Bacillus sphaericus* complex. The modern *B. sphaericus* complex is incompletely known, so the "ancient" sequence obtained could be that of a modern, but previously unidentified, bacterium because these bacteria often are isolated from the soil (Yousten and Rippere 1997).

Other claims of amplifying ancient DNA have been disproved. For example, the mitochondrial cytochrome *b* sequence of an 80-million-year-old dinosaur from the Upper Cretaceous in Utah was later discovered to be, most probably, of human origin (Hedges and Schweitzer 1995). Likewise, a 20-million-year-old magnolia leaf produced sequences that were similar to those of modern magnolias. The authenticity of the magnolia sequences was cast into doubt because they were exposed to water and oxygen during preservation and DNA is especially vulnerable to degradation under such conditions.

The jury is out on the authenticity of ancient DNA in insects embedded in amber. Fortunately, the scientific criteria for resolving the controversy are now more clear.

The most common ancient DNA analyzed is usually mitochondrial DNA because it is so abundant. However, this abundance makes it easy to contaminate the ancient sample with modern mtDNA. The amplification of ancient DNA remains highly controversial because technical difficulties are enormous (Rollo 1998, Cooper and Poinar 2000).

DNA decays spontaneously, mainly through hydrolysis and oxidation. Hydrolysis causes deamination of the nucleotide bases and cleavage of base-sugar bonds, creating baseless sites. Deamination of cytosine to uracil and depurination (loss of purines adenine and guanine) are two types of hydrolytic damage. Baseless sites weaken the DNA, causing breaks that fragment the DNA into smaller and smaller pieces. Oxidation leads to chemical modification of bases and destruction of the ring structure of base and sugar residues (Austin et al. 1997a). As a result, it is almost always impossible to obtain long amplification products from ancient DNA (Handt et al. 1994). It is possible to use overlapping primer pairs if longer sequences are needed, but there usually is an inverse relationship between efficiency and length of the PCR products. When such an inverse relationship is *not* seen, the amplification product often turns out to be due to contamination (Handt et al. 1994, Hofreiter et al. 2001).

PCR products from ancient DNA often are "scrambled." This is due to the phenomenon called "jumping PCR," which occurs when the DNA polymerase reaches a template position which carries either a lesion or a strand break that stops the polymerase (Handt et al. 1994). The partially extended primer can anneal to another template fragment in the next cycle and be extended up to another damaged site. Thus, *in vitro* recombination (jumping) can take place until the whole stretch encompassed by the two primers is synthesized and the amplification enters the exponential part of the PCR (Handt et al. 1994). This phenomenon makes it essential that cloning and sequencing of multiple clones be carried out to eliminate this form of error.

Most archeological and paleontological specimens contain DNA from exogenous sources such as bacteria and fungi, as well as contaminating DNA from contemporary humans (Poinar and Stankiewicz 1999). Aspects of burial conditions seem to be important in DNA preservation, especially low temperature during burial (Poinar and Stankiewicz 1999). The oldest DNA sequences reported and confirmed in other laboratories come from the remains of a wooly mammoth found in the Siberian permafrost; these sequences are "only" 50,000 years old—not millions of years old (Poinar and Stankiewicz 1999).

Theoretical calculations and empirical observations suggest DNA should only be able to survive, in a highly fragmented and chemically modified form, for 50,000 to 100,000 years (Austin et al. 1997a, Rollo 1998, Hofreiter et al. 2001). Because only tiny amounts of DNA usually can be extracted from an archeological specimen, stringent precautions and multiple controls are required to avoid accidental contamination with modern DNA.

A methodology to deal with ancient specimens has been proposed that includes careful selection of well-preserved specimens, choice of tissue samples that are likely to have the best DNA preservation, and surface sterilization to eliminate surface contamination. The operations should be carried out in a laboratory dedicated to work on ancient specimens, and work on ancient DNA should be separated from that on modern DNA (Austin et al. 1997a, Cooper and Poinar 2000, Hofreiter et al. 2001). Most importantly, multiple negative controls should be performed during DNA extraction and PCR setup, although a lack of positives in the negative controls is not definitive proof of authentic ancient DNA.

Another crucial step is the authentication of the results. Putatively ancient DNA sequences should be obtained from different extractions of the same sample and from different tissue

samples from different specimens (Austin et al. 1997a, Cooper and Poinar 2000). The ultimate test of authenticity should be independent replication in two separate laboratories (Rollo 1998). So far, this type of replication has not been achieved for DNA from amber-preserved arthropod specimens (Austin et al. 1997b, Walden and Robertson 1997, Gutierrez and Marin 1998).

8.5.2. Amplifying Old DNA

Amplification of old DNA from museum specimens is less difficult and less controversial (Paabo 1990, 1991, Jackson et al. 1991, Cano et al. 1993a,b, Townson et al. 1999). DNA from pathogens contained within museum specimens of arthropods can be amplified by the PCR. For example, Lyme disease spirochete (*Borrelia burgdorferi*) DNA extracted from the midgut of ticks (*Ixodes dammini*) stored for 50 years in 70% EtOH could be amplified by the PCR (Persing et al. 1990). Individual tick specimens were removed from the EtOH with flame-sterilized forceps and air-dried on filter paper disks for 5 min. Then, 200 μl of 0.5-mm glass beads were incubated in 1 ml of 1% bovine serum albumin in distilled water at 37°C for 30 min and then washed twice in 1 ml of distilled water. Ticks were placed whole into 0.5-ml microcentrifuge tubes containing a slurry (20 μl) of the treated glass beads. Specimens were crushed into the beads with a disposable plastic dowel for 30 to 45 sec to liberate the midgut contents, and 25 μl of PCR buffer was added. Samples were boiled for 5 min, then cooled on ice; 5-μl portions of the supernatant fluid were used for the PCR.

A simpler protocol was used by Azad et al. (1990) to determine whether individual ticks or fleas were infected with rickettsia. Individual ticks or fleas were placed in 100 μl of brain heart infusion broth and boiled for 10 min. The PCR was carried out with 10 μl of the suspension. Because the PCR can be applied to frozen or formalin-fixed tissues, dried museum specimens, and alcohol-preserved specimens, PCR can reduce the potential dangers involved in maintaining and transporting live infectious disease vectors from the field. In addition, detection of pathogens by the PCR is significantly more sensitive than by ELISA (Azad et al. 1990).

Dried, pinned specimens of the *Anopheles gambiae* mosquito complex, ranging in age from 15 to 93 years, were tested to determine if ribosomal DNA could be amplified by the PCR (Townson et al. 1999). Most of the specimens yielded amplifiable DNA from entire abdomens, but extractions from single hind legs from these old, dried specimens were unsuccessful. By contrast, single legs from a fresh specimen produced sufficient DNA to yield a PCR product. Note, however, that ribosomal genes are present in high copy numbers.

The PCR has been used to amplify DNA from tissues preserved in formalin followed by paraffin embedding. Specimens up to 40 years old have yielded DNA up to 800 bp in length (Wright and Manos 1990). The integrity of the DNA and the duration of fixation affect the length of the product that can be amplified.

8.5.3. Amplifying RNA

The PCR can be used to amplify messenger RNA sequences from complementary DNA (Kawasaki 1990). This allows analysis of gene expression during development, quantitation of mRNA from specific tissues, rearrangements of DNA during cell differentiation, and RNA processing.

8.5.4. Analysis of mRNA Polyadenylation

Measurement of poly(A) tail length is important when studying mRNA stability. A simple PCR-based method has been developed (Eguchi and Eguchi 2000).

8.5.5. Cloning a Gene

The PCR can generate microgram quantities of a specific DNA fragment, and these can be cloned (Scharf 1990), although many products of the PCR are "recalcitrant to cloning" (Sambrook and Russell 2001). One reason is that several of the DNA polymerases used in the PCR have the ability to add a single, unpaired nucleotide at the 3′ end of the PCR product (terminal transferase activity). The nucleotide added depends both on the adjacent base and on the particular polymerase used; for example, when the 3′ terminal base of the template DNA is cytosine, *Taq* will add an adenine (A) to the end of the completed PCR product (Sambrook and Russell 2001).

One solution to this problem is to use the 3′ to 5′ exonuclease activity of bacteriophage T4 DNA polymerase or *Pfu* DNA polymerase to "polish" the ends of the PCR products that contain the added bases; the polished DNA fragments can then be phosphorylated by T4 kinase and cloned into a blunt-ended dephosphorylated vector (Costa and Weiner 1994, Costa et al. 1994). Unfortunately, blunt-ended cloning is notoriously inefficient (10- to 100-fold less efficient than cloning with DNA fragments with cohesive termini). Furthermore, blunt-ended cloning allows no opportunity to direct the orientation of the fragment within the vector.

A second reason for potential difficulties in cloning a PCR product is that *Taq* (and perhaps other polymerases) can survive extraction with various methods used to purify the PCR products (Bennett and Molenaar 1994, Sambrook and Russell 2001). The residual polymerase and dNTPs may make it difficult to tailor the ends of the amplified DNA for cloning.

Currently, the most popular and efficient method for cloning PCR products involves a method that relies on ligation of cohesive ends of the PCR product with a vector (Sambrook and Russell 2001). *Taq* DNA polymerase typically adds an A at the 3′ end of the product. If a plasmid vector is used with a protruding 3′ T residue at each of its ends, cloning is more efficient because the T and A can base pair. It is perhaps surprising that pairing of single bases is sufficient to increase efficiency of cloning, but it is estimated to be approximately 50-fold more efficient than blunt-ended cloning.

Restriction sites frequently are inserted at the 5′ end of each primer so that the amplified DNA can be cloned directly into a vector after digestion of the amplified DNA (Kaufman and Evans 1990). Because the restriction sites can be the same or different in the two primers, the researcher can tailor the ends of the PCR product to the specific vectors required for the project (Sambrook and Russell 2001). Various commercial kits allow direct cloning of PCR products.

The isolation of a gene requires some prior knowledge of the gene sequence (Clackson et al. 1991). If a probe (primer) is available from another species, genomic DNA can be screened by the PCR using standard or degenerate primers (McPherson et al. 1991, Clackson et al. 1991). The success of this approach was illustrated by the cloning of a sodium channel gene from *Drosophila* and the house fly, *Musca domestica*, using degenerate primers (Knipple et al. 1991). Several vertebrate sodium channel genes had been cloned, and comparisons of the inferred amino acid sequences of the alpha subunits of sodium channels from rat brain and rat skeletal muscle to that of the electric eel revealed a 70%

homology when conservative substitutions were taken into account. Two sodium channel genes (*para* and DSC1) cloned from *Drosophila* were homologous to the vertebrate sodium channel genes. Using this information, it was possible to generate DNA primers to isolate a segment of the gene homologous to *para* from the house fly.

The PCR was performed on genomic house-fly DNA using degenerate primers. The 5′-end primer consisted of a 256-fold degenerate sequence 20 nt long. The 3′-end primer consisted of a 64-fold degenerate sequence 21 nt long. Both had additional sequences appended to their 5′ ends to provide a *Hin*dIII and *Xba*I restriction enzyme recognition sequence, respectively, to facilitate cloning the amplification products. The PCR product was 104 bp long, consisting of 87 bp of coding sequence plus the flanking sequences attached to the 5′ ends of the primers. To confirm that the PCR-generated DNA was derived from the house fly, amplified DNA was labeled with ^{32}P and used as a probe of genomic Southern blots containing digests of house fly, *Drosophila*, and mouse DNA. Because the only specific hybridization signal after high-stringency washing was to the house-fly DNA, the amplified DNA was not an artifact or contaminant. The PCR products were cloned and sequenced, and the sequence isolated from the house fly differed from that of *Drosophila* at only 16 nucleotides (81.6% similarity). The substitutions, primarily in the third codon, had no effect on the amino acid sequence.

Doyle and Knipple (1991) subsequently used the same degenerate (mixed sequence) primers to amplify DNA from seven insects and an arachnid, including the tobacco budworm *Heliothis virescens*, the mosquito *Aedes aegypti*, the diamondback moth *Plutella xylostella*, the gypsy moth *Lymantria dispar*, the cabbage looper *Trichoplusia ni*, the Colorado potato beetle *Leptinotarsa decemlineata*, the American cockroach *Periplaneta americana*, and the two-spotted spider mite *Tetranychus urticae*. Following amplification, the PCR products were sequenced, and only 5 of 60 clones were not derived from *para* homologs. This study, and others, suggest that degenerate primers derived from conserved segments of characterized *D. melanogaster* genes can be used to clone genes from a diverse array of arthropods. Interestingly, intraspecific polymorphisms were found in the sequence from three moth species, which could reflect the presence of duplicated genes or allelic variants in the populations. Doyle and Knipple (1991) suggested that the *para* gene could be used to analyze insect populations for pesticide resistance.

Similarly, primers for the conserved *Actin* gene(s) in insects were used to clone these genes from the predatory mite *Metaseiulus occidentalis*, despite the long evolutionary separation of insects and mites (Hoy et al. 2000). The rich source of sequence information in GenBank for the complete *Drosophila* genome makes this approach increasingly feasible.

8.5.6. Detecting Gene Amplification

Sometimes it is important to determine whether a gene has been amplified (increased in copy number), leading to increased levels of gene product. Some insects are resistant to pesticides because of amplification of esterase genes. A method called comparative PCR can be used to detect gene amplification (Brass et al. 1998).

8.5.7. Detecting Methylation of DNA

Genomic imprinting is often due to DNA methylation at several sites in the genome. A methylation-specific PCR assay can be used to detect methylation of specific genes more quickly than the use of Southern blot assays (Kubota et al. 1997).

8.5.8. Detecting Pathogens in Vector Arthropods

Arthropod vectors such as ticks, fleas, and mosquitoes are involved in maintaining and transmitting (vectoring) pathogens to humans and other vertebrates. Aphids and leafhoppers transmit (vector) viruses and mycoplasma to plants.

The detection of pathogenic microorganisms within vector arthropods is important in conducting epidemiological studies and developing control strategies. A number of antigen-detection techniques have been developed, including direct or indirect immunofluorescence tests and enzyme-linked immunosorbent assays (ELISA) using polyclonal or monoclonal antibodies. Other techniques involve recovery of the microorganisms from vectors by culture in embryonated eggs or tissue culture cells or by experimental infections in laboratory animals. The recovery of pathogenic microorganisms by these methods requires either live or properly frozen specimens. These techniques are expensive and time consuming, and they may involve dissection and preparation of specimens from live arthropods, which can be hazardous (Barker 1994).

The PCR offers another approach to detecting and identifying pathogenic microorganisms if sequence information is available to design appropriate primers (Wise and Weaver 1991, Higgins and Azad 1995). The PCR can be carried out with material from dead specimens, is more sensitive than most immunological techniques, and is more rapid. For example, using primers that amplify a 434-bp fragment of a protein antigen from fleas and ticks infected with *Rickettsia rickettsii* have been identified (Azad et al. 1990). Malarial DNA has been detected in both infected blood and individual mosquitoes (Schriefer et al. 1991). As few as three *Leishmania* parasites could be detected in infected sand flies using seminested PCR (Aransay et al. 2000). Trypanosome infections could be detected in wild tsetse flies in Cameroon (Morlais et al. 1998). The heartwater fever pathogen (*Cowdria ruminantium*) could be detected in vector ticks (*Amblyomma*) with high levels of specificity when 10^7 to 10^4 organisms were present. The reliability of the assay dropped when ticks had only 10^3 to 10^2 organisms, which highlights the need to conduct quantitative analyses for sensitivity before employing PCR assays in epidemiological studies (Peter et al. 2000).

West Nile virus was detected in human clinical specimens, field-collected mosquitoes, and bird samples by a TaqMan reverse transcriptase PCR assay (Lanciotti et al. 2000). This rapid, specific, and sensitive assay can be used in the diagnostic laboratory for testing humans and as a tool for conducting surveillance of West Nile virus in mosquitoes and birds in the field (Anderson et al. 1999). Sequencing of the West Nile virus causing encephalitis in the northeastern United States indicated the virus was most closely related to a virus isolated in Israel in 1998 (Lanciotti et al. 1999).

A quantitative PCR protocol was used to assay densities of the plague bacteria *Yersinia pestis* in fleas and mice. The assays indicated fleas needed $\sim 10^6$ bacteria to be able to transmit the bacteria to mice (Engelthaler et al. 2000).

Random primers (hexanucleotides) were used to develop primers for a multiplex reverse transcriptase PCR to detect five potato viruses and a viroid in aphids, leaves, and potato tubers (Lie and Singh 2001).

8.5.9. Detecting Pesticide Resistance

The malaria vectors *Anopheles gambiae* and *A. arabiensis* were screened for permethrin resistance (nerve insensitivity, *kdr*-type) (Brooke et al. 1999). The results indicated that one of the populations was resistant to permethrin through a different biochemical method,

indicating that both PCR and bioassay data should be obtained for monitoring resistance allele frequencies and the operational mode of insecticide resistance.

8.5.10. Developmental Biology

It is possible to detect the presence of specific mRNAs in tissues or cells by reverse transcription and DNA amplification by the PCR.

8.5.11. Engineering DNA

DNA can be engineered in several ways by the PCR. Sequences can be added to the 5′ end of primers. Such sequence changes are readily accepted, even though these add-ons do not base pair with the template DNA. The DNA being synthesized contains the add-on because the primers are incorporated in the synthesized DNA fragment. For example, it is possible to add a restriction-site sequence to DNA being amplified by the PCR by attaching the restriction-site sequence to the primers (Figure 8.7). Such restriction sites facilitate subsequent manipulations of the final PCR product.

The T7 promoter located at the 5′ end of one primer can be added to PCR products. This promoter allows RNA copies to be generated from the DNA synthesized in the PCR reaction. Although the add-on sequences in the primers don't base pair to the template DNA, in most cases they have little effect on the specificity or efficiency of the amplification. Specificity is apparently imparted most significantly by the 3′ end of the primer.

One PCR product strand or the other or both can be tagged with a radioactive, biotin, or fluorescent label (Chehab and Kan 1989, Mertz and Rashtchian 1994). DNA sequences can also be altered at any position by modifying primers so that substitutions, additions, or deletions are made in the amplified DNA.

8.5.12. Evaluating Efficacy of Disease Control

In regions of the world where malaria is endemic, the use of bednets impregnated with a synthetic pyrethroid insecticide has been proposed as a method to control the mosquito vector (Gokool et al. 1992). The PCR was used to determine whether pesticide-treated bednets prevent mosquito transmission of malaria. PCR was used to amplify and fingerprint the human DNA contained within a mosquito blood meal. DNA fingerprints also were obtained from the blood of individuals sleeping under the bednets and compared to the fingerprints from the mosquitoes (Gokool et al. 1992). The many hypervariable regions of the human

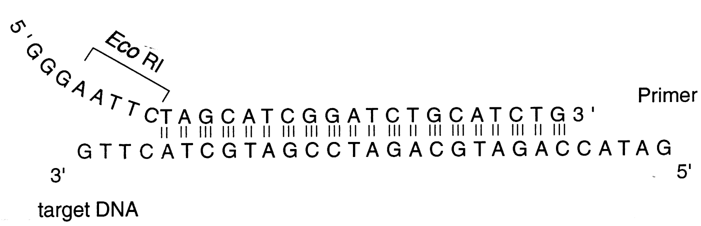


Figure 8.7. A 5′ add-on of a restriction site sequence (*Eco*RI) to a primer, which is annealed to a target DNA sequence. Although the add-on does not specifically match the template DNA, this does not significantly affect the PCR. The extra bases that are added 5′ to the *Eco*RI site ensure that the efficiency of the restriction enzyme cleavage is maintained.

genome produce individual-specific patterns of DNA fragments, and the banding patterns obtained indicated that few mosquitoes had fed on individuals protected by treated bednets.

8.5.13. Evolutionary Analyses

Analysis of evolution involves reconstruction of phylogenies and analysis of population genetics (see Chapter 13 for further details).

The relative ease and simplicity of the PCR enhances molecular studies of evolution. The PCR makes it possible to directly sequence amplified gene fragments from individuals and populations, increasing the resolving power and phylogenetic range of comparative studies. For most applications of the PCR, it is necessary to know a sequence to synthesize primers. However, by choosing sequences that are highly conserved among widely divergent species, it is possible to design "universal primers" to amplify a particular nuclear or organelle gene fragment from many members of a major taxonomic group. This allows comparisons of sequences from classes or phyla for taxonomic work as well as enhancing population studies that involve identifying individuals and biotypes.

Universal primers have been developed for a number of nuclear and mitochondrial genes. For example, primers that amplify a region of approximately 515 bp of the 18S rDNA from many fungi, protozoa, algae, plants, and animals are available. The primers are based on conserved sequences among the 18S rDNA, but they do not amplify bacterial or mitochondrial rRNA genes.

Mitochondrial gene sequences are useful for many problems in evolutionary and population biology (Kapsa et al. 1997). Primers have been developed that allow a number of different gene fragments to be amplified from different insect orders (Simon et al. 1994, Kambhampati and Smith 1995). It even may be possible to amplify complete insect mitochondrial genomes in "two easy pieces" (Roehrdanz 1995). Because mitochondria are inherited maternally, sequence analyses of mitochondrial DNA allow construction of maternal phylogenies. Mitochondrial DNA evolves at a higher rate than nuclear DNA because mitochondria lack a proofreading function to correct errors in DNA synthesis, which makes mitochondria especially useful for analyses of closely related populations or species.

8.5.14. Sequencing DNA

Both the Maxam–Gilbert and Sanger sequencing methods have been modified to sequence DNA amplified by the PCR (Ausubel et al. 1991, Ellingboe and Gyllensten 1992, Kocher 1992, Olsen et al. 1993, Rao 1994a,b). Both methods permit the rapid determination of sequences without the need to construct a library or screen the library for the gene(s) of interest. Sequencing of PCR products can be either direct (Landweber and Kreitman 1995) or after amplification and cloning (Olsen et al. 1993, Sambrook and Russell 2001). Kits are available that aid in sequencing by various methods.

8.5.14.1. Cycle Sequencing

Cycle sequencing is a method in which asymmetric PCR is used to generate a ss DNA template for sequencing by the Sanger dideoxy-chain termination method (Sambrook and Russell 2001).

In cycle sequencing, four separate amplification reactions are set up, each containing the same primer and a different chain-terminating ddNTP. Two cycling programs are used;

during the first, the reaction mixtures are amplified for 15 to 40 rounds by denaturation of the template DNA, annealing of the ^{32}P-labeled sequencing primer to the target, and termination of the extended strand by incorporation of a ddNTP. The result is a hybrid molecule that is partially double-stranded and consists of the full-length template strand and its complementary chain-terminated product. This product is denatured during the first step of the next cycle, resulting in the template strand becoming available for another round of priming, extension, and termination. Cycle sequencing thus produces product in a linear fashion.

In the second program, the annealing step is omitted so that no further extension of primers is possible. Instead, the second segment provides an opportunity to further extend the reaction products that were not terminated by incorporation of a ddNTP during the initial rounds of the PCR. The radiolabeled products are displayed on a denaturing polyacrylamide gel and detected by autoradiography.

Advantages of cycle sequencing include the fact that it works with both ds and ss DNA templates, which means that it is unnecessary to subclone the DNA into M13 or other ss vectors. Cycle sequencing also requires only tiny amounts of template, which can be obtained from a single plaque or colony or from a PCR product purified from a gel. Other advantages are that it can be set up in either microtiter plates or microfuge tubes, it can be used with commercially available robotic workstations, and it can be modified to obtain the sequence of each strand of a ds DNA template (Sambrook and Russell 2001).

Disadvantages of cycle sequencing for small laboratories with limited equipment are that cycle sequencing with radiolabeled primers requires dedicating the equipment (cycler, gel equipment) in order to avoid contamination with radioactivity. Also, cycle sequencing requires very clean template DNA without agarose, salts, or proteins that could cause premature termination of the DNA polymerase. These impurities can lead to a high level of false bands and “empty” bands.

8.5.14.2. Direct Sequencing

Direct sequencing of ds PCR products without cloning into ss sequencing vectors should, in theory, save time and effort. Unfortunately, it can be “unreliable in practice” (Sambrook and Russell 2001).

Direct sequencing of PCR products can only be successful if the PCR was optimized to reduce mispriming and the PCR product was cleaned of residual primers, DNA polymerase, unused dNTPs, and nonspecific copies of the original template (Sambrook and Russell 2001). Effective cleaning of the PCR product may be achieved through spun-column chromatography or centrifugal ultrafiltration to remove residual primers and unused dNTPs. Residual thermostable DNA polymerase and residual dNTPs may “befoul chain-termination sequencing reactions” catalyzed by other thermostable enzymes, such as Sequenase and AmpliTaq, unless eliminated (Sambrook and Russell 2001). Elimination of nonspecific PCR products can be achieved by running the DNA on a low-melting-point agarose gel and separating the band of the appropriate size from the gel.

Direct sequencing of DNA amplified by a PCR protocol that uses only a single primer (such as RAPD-PCR) requires a different procedure (Iizuka et al. 1996). If ss DNAs first are isolated by polyacrylamide gel electrophoresis, sequencing can be conducted on the single strands using the corresponding single primers.

Direct sequencing of PCR-amplified DNA is not appropriate if the starting sample contained one normal allele and one deleted allele because the deleted allele will be masked. Likewise, if the sample contains multiple alleles, direct sequencing would result in a

composite sequence ladder, making it impossible to decipher the sequence of individual alleles (Rao 1994a). Under these circumstances, the amplified DNA should be cloned and the sequences of a number of clones should be determined to resolve the sequences of individual alleles.

Direct sequencing of PCR-amplified DNA is inappropriate if nonspecific products were produced. Thus, direct sequencing requires good PCR conditions: quality DNA template, highly specific primers, high annealing temperature, initiating the PCR by the "hot-start" method, and performing a rapid cycling protocol for as few cycles as possible using an adequate amount of target DNA (Rao 1994a).

8.6. Concluding Remarks

The speed, specificity, versatility, and sensitivity of the PCR has had a significant effect on genetics, immunology, forensic science, evolutionary biology, systematics, ecology, and population biology (Arnheim and Erlich 1992, Dieffenbach 1995). The PCR has revolutionized the way in which much of our research is conducted. End users and commercial developers continue to develop sophisticated techniques for integrating the PCR into more applications, including diagnostic assays using the PCR and electrophoretic analysis on a microchip (Waters et al. 1998).

References Cited

Agusti, N., M. C. de Vicente, and R. Gabarra (2000). Developing SCAR markers to study predation on *Trialeurodes vaporariorum*. *Insect Mol. Biol.* **9**: 263–268.

Aljanabi, S. M., and I. Martinez (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* **25**: 4692–4693.

Al-Soud, W. A., and P. Radstrom (1998). Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Appl. Env. Microbiol.* **64**: 3748–3753.

Anderson, J. F., T. G. Andreadis, C. R. Vossbrinck, S. Tirrell, E. M. Wakem, R. A. French, A. E. Garmendia, and H. J. Van Kruiningen (1999). Isolation of West Nile virus from mosquitoes, crows and a Cooper's hawk in Connecticut. *Science* **286**: 2331–2333.

Aransay, A. M., E. Scoulica, and Y. Tselentis (2000). Detection and identification of *Leishmania* DNA within naturally infected sand flies by seminested PCR on minicircle kinetoplastic DNA. *Appl. Env. Microbiol.* **66**: 1933–1938.

Arnheim, N., and H. Erlich (1992). Polymerase chain reaction strategy. *Annu. Rev. Biochem.* **61**: 131–156.

Arnheim, N., T. White, and W. E. Rainey (1990). Application of PCR: organismal and population biology. *BioScience* **40**: 174–182.

Asgari, S., J. R. E. Hardy, R. G. Sinclair, and B. D. Cooke (1998). Field evidence for mechanical transmission of rabbit haemorrhagic disease virus (RHDV) by flies (Diptera: Calliphoridae) among wild rabbits in Australia. *Virus Res.* **54**: 123–132.

Austin, A. D., and N. Dillon (1997). Extraction and PCR of DNA from parasitoid wasps that have been chemically dried. *Austr. J. Entomol.* **36**: 241–244.

Austin, J. J., A. B. Smith, and R. H. Thomas (1997a). Palaeontology in a molecular world: the search for authentic ancient DNA. *Trends Evol. Ecol.* **12**: 303–306.

Austin, J. J., A. J. Ross, A. B. Smith, R. A. Fortey, and R. H. Thomas (1997b). Problems of reproducibility—does geologically ancient DNA survive in amber-preserved insects? *Proc. R. Soc. Lond.* B **264**: 467–474.

Ausubel, F. M., R. Brent, R. E. Kingston, D. M. Moore, J. G. Seldman, J. A. Smith, and K. Struhl, Eds. (1991). The polymerase chain reaction. *Current Protocols in Molecular Biology*. Greene Publ., New York.

Ayliffe, M. A., G. J. Lawrence, J. G. Ellis, and A. J. Pryor (1994). Heteroduplex molecules formed between allelic sequences cause nonparental RAPD bands. *Nucleic Acids Res.* **22**: 1632–1636.

Ayyadevara, S., J. J. Thaden, and R. J. Shmookler Reis (2000). Discrimination of primer 3′-nucleotide mismatch by *Taq* DNA polymerase during polymerase chain reaction. *Anal. Biochem.* **284**: 11–18.

Azad, A. F., L. Webb, M. Carl, and G. A. Dasch (1990). Detection of rickettsiae in arthropod vectors by DNA amplification using the polymerase chain reaction, pp. 557–563. In: *Rickettsiology: Current Issues and Perspectives*, Vol. 590, Ann. N.Y. Acad. Sci.

Baldrich, E., N. Xamena, and O. Cabre (1999). Overcoming false negatives due to the genomic context in polymerase chain reaction amplification. *J. Biochem. Biophys. Methods* **40**: 45–48.

Barker, R. H., Jr. (1994). Use of PCR in the field. *Parasitol. Today* **10**: 117–119.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. USA* **91**: 2216–2220.

Beckenbach, A. T. (1995). Age of bacteria from amber. *Science* **270**: 2015–2016.

Bennett, B. L., and A. J. Molenaar (1994). Cloning of PCR products can be inhibited by *Taq* polymerase carryover. *BioTechniques* **16**: 32–33.

Brass, N., D. Heckel, and E. Meese (1998). Comparative PCR: An improved method to detect gene amplification. *BioTechniques* **24**: 22–26.

Brooke, B. D., R. H. Hunt, L. L. Koekemoer, J. Dossou-Yovo, and M. Coetzee (1999). Evaluation of a polymerase chain reaction assay for detection of pyrethroid insecticide resistance in the malaria vector species of the *Anopheles gambiae* complex. *J. Amer. Mosq. Contr. Assoc.* **15**: 565–568.

Brownie, J., S. Shawcross, J. Theaker, D. Whitcombe, R. Ferrie, C. Newton, and S. Little (1997). The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res.* **25**: 3235–3241.

Burt, A., D. A. Carter, T. J. White, and J. W. Taylor (1994). DNA sequencing with arbitrary primer pairs. *Mol. Ecol.* **3**: 523–535.

Cano, R. J., and M. K. Borucki (1995). Revival and identification of bacterial spores in 25- to 40-million-year-old Dominican amber. *Science* **268**: 1060–1063.

Cano, R. J., H. N. Poinar, N. J. Pieniazek, A. Acra, and G. O. Poinar, Jr. (1993a). Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature* **363**: 536–538.

Cano, R. J., H. N. Poinar, D. W. Roubik, and G. O. Poinar, Jr. (1993b). Enzymatic amplification and nucleotide sequencing of portions of the 18s rRNA gene of the bee *Proplebeia dominicana* (Apidae: Hymenoptera) isolated from 25–40 million year old Dominican amber. *Med. Sci. Res.* **20**: 619–622.

Carbonari, M., D. Sbarigia, M. Cibati, and M. Fiorilli (1993). Optimization of PCR performance. *Trends Genet.* **9**:42–43.

Chan, C. K., D. J. Petersen, and T. C. Vrain (1999). DNA fingerprinting of single aphid embryos by random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR). *Can. Entomol*. **131**: 229–230.

Chehab, F. F., and Y. W. Kan (1989). Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. *Proc. Natl. Acad. Sci. USA* **86**: 9178–9182.

Cheng, S. (1995). Longer PCR amplifications, pp. 313–324. In: *PCR Strategies*. M. A. Innis, D. H. Gelfand, and J. J. Sninsky, Eds. Academic Press, San Diego.

Cheng, S., and L. A. Kolmodin (1998). XL PCR amplification of long targets from genomic DNA, pp. 17–29. In: *Methods in Molecular Biology*. Vol. 67: *PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering*, B. A. White, Ed. Humana Press, Totowa, NJ.

Cheng, S., C. Fockler, W. M. Barnes, and R. Higuchi (1994a). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci. USA* **91**: 5695–5699.

Cheng, S., S. Y. Chang, P. Gravitt, and R. Respess (1994b). Long PCR. *Nature* **369**: 684–685.

Chevet, E., G. Lemaitre, and M. D. Katinka (1995). Low concentrations of tetramethylammonium chloride increase yield and specificity of PCR. *Nucleic Acids Res.* **23**: 3343–3344.

Clackson, T., D. Gussow, and P. T. Jones (1991). General application of PCR to gene cloning and manipulation, pp. 187–214. In: *PCR. A Practical Approach*. M. J. McPherson, P. Quirke and G. R. Taylor, Eds. IRL Press, Oxford.

Collasius, M., H. Puchta, S. Schlenker, and G. Valet (1991). Analysis of unknown DNA sequences by polymerase chain reaction (PCR) using a single specific primer and a standardized adaptor. *J. Virol. Methods* **32**: 115–119.

Comes, A. M., J. F. Humbert, and F. Laurent (1997). Rapid cloning of PCR-derived RAPD probes. *BioTechniques* **23**: 211–212.

Cooper, A., and H. N. Poinar (2000). Ancient DNA: do it right or not at all. *Science* **289**: 1139.

Costa, G. L., and M. P. Weiner (1994). Polishing with T4 or *Pfu* polymerase increases the efficiency of cloning of PCR fragments. *Nucleic Acids Res.* **22**: 2423.

Costa, G. L., A. Grafsky, and M. P. Weiner (1994). Cloning and analysis of PCR-generated DNA fragments. *PCR Methods Appl.* **2**: 338–345.

Czerny, T. (1996). High primer concentration improves PCR amplification from random pools. *Nucleic Acids Res.* **24**: 985–986.

DeSalle, R., J. Gatesy, W. Wheeler, and D. Grimaldi (1992). DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science* **257**: 1933–1936.

Desmarais, E., I. Lanneluc, and J. Lagnel (1998). Direct amplification of length polymorphisms (DALP), or how to get and characterize new genetic markers in many species. *Nucleic Acids Res.* **26**: 1458–1465.

Dick, M., D. M. Bridge, W. C. Wheeler, and R. DeSalle (1993). Collection and storage of invertebrate samples. *Methods Enzymol.* **224**: 51–65.

Dieffenbach, C. W., Ed. (1995). *PCR Primer: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Doyle, K. E., and D.C. Knipple (1991). PCR-based phylogenetic walking: isolation of *para*-homologous sodium channel gene sequences from seven insect species and an arachnid. *Insect Biochem.* **21**: 689–696.

Dwyer, D. E., and N. Saksena (1992). Failure of ultra-violet irradiation and autoclaving to eliminate PCR contamination. *Mol. Cell. Probes* **6**: 87–88.

Ebert, P. R. (1996). Isolation of megabase-sized DNA fragments from the honey bee (*Apis mellifera*) genome. *J. Apic. Res.* **35**: 124–125.

Eckert, K. A., and T. A. Kunkel (1991). The fidelity of DNA polymerases used in the polymerase chain reactions, pp. 225–245. In: *PCR. A Practical Approach*, M. J. McPherson, P. Quirke and G. R. Taylor, Eds., IRL, Oxford.

Edwards, O. W., and M. A. Hoy (1993). Polymorphisms in two parasitoids detected using random amplified polymorphic DNA (RAPD) PCR. *Biol. Control* **3**: 243–257.

Edwards, O. W., and M. A. Hoy (1994). Monitoring laboratory and field biotypes of the walnut aphid parasite, *Trioxys pallidus*, in population cages using RAPD-PCR. *Biocontr. Sci. Technol.* **5**: 313–327.

Eguchi, Y., and T. Eguchi (2000). PCR-based method for the rapid analysis of mRNA polyadenylation. *Biotechnol. Lett.* **22**: 1583–1586.

Ellingboe, J., and U. B. Gyllensten, Eds. (1992). *The PCR Technique: DNA Sequencing*. Eaton Publ., Natick, MA.

Ellsworth, D. L., K. D. Rittenhouse, and R. L. Honeycutt (1993). Artifactual variation in Randomly Amplified Polymorphic DNA banding patterns. *BioTechniques* **14**: 214–217.

Engelke, D. R., A. Krikos, M. E. Bruck, and D. Ginsburg (1990). Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. *Anal. Biochem.* **191**: 396–500.

Engelthaler, D. M., B. J. Hinnebusch, C. M. Rittner, and K. L. Gage (2000). Quantitative competitive PCR as a technique for exploring flea–*Yersina pestis* dynamics. *Am. J. Trop. Med. Hyg.* **62**: 552–560.

Erlich, H. A., Ed. (1989). *PCR Technology, Principles and Applications of DNA Amplification*, Stockton Press, New York.

Erlich, H. A., and N. Arnheim (1992). Genetic analysis using the polymerase chain reaction. *Annu. Rev. Genetics* **26**: 479–506.

Erlich, H. A, R. Gibbs, and H. H. Kazazian, Jr., Eds. (1989). *Polymerase Chain Reaction*. *Current Communications in Molecular Biology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Erlich, H. A., D. Gelfand, and J. J. Sninsky (1991). Recent advances in the polymerase chain reaction. *Science* **252**: 1643–1651.

Espinasa, L., and R. Borowsky (1998). Evolutionary divergence of AP-PCR (RAPD) patterns. *Mol. Biol. Evol.* **15**: 408–414.

Foley, K. P., M. W. Leonard, and J. D. Engel (1993). Quantitation of RNA using the polymerase chain reaction. *Trends Genet*. **9**: 380–385.

Forneck, A., M. A. Walker, and R. Blaich (2000). Genetic structure of an introduced pest, grape phylloxera (*Daktulosphaira vitifoliae* Fitch), in Europe. *Genome* **43**: 669–678.

Frothingham, R., R. B. Blitchington, D. H. Lee, R. C. Greene, and K. H. Wilson (1992). UV absorption complicates PCR decontamination. *BioTechniques* **13**: 208–210.

Fukatsu, T. (1999). Acetone preservation: a practical technique for molecular analysis. *Mol. Ecol.* **8**: 1935–1945.

Fukatsu, T., and H. Ishikawa (1994). Differentiation of aphid clones by arbitrarily primed polymerase chain reaction (AP-PCR) DNA fingerprinting. *Mol. Ecol.* **3**: 187–192.

Gerken, T., J. Kurtz, K. P. Sauer, and T. Lubjuhn (1998). DNA preparation and efficient microsatellite analysis from insect hemolymph. *Electrophoresis* **19**: 3069–3070.

Gokool, S. D., F. Smith, and C.F. Curtis (1992). The use of PCR to help quantify the protection provided by impregnated bednets. *Parasitol. Today* **8**: 347–350.

Goldenberger, D., I. Perschil, M. Ritzler, and M. Altwegg (1995). A simple "universal" DNA extraction procedure using SDS and proteinase K is compatible with direct PCR amplification. *PCR Methods Appl*. **4**: 368–370.

Golenberg, E. M., A. Bickel, and P. Weihs (1996). Effect of highly fragmented DNA on PCR. *Nucleic Acids Res.* **24**: 5026–5033.

Goodman, M. F. (1995). DNA polymerase fidelity: misinsertions and mismatched extensions, pp. 17–31. In: *PCR Strategies*, Academic Press, San Diego.

Grevelding, C. G., A. Kampkotter, M. Hollmann, U. Schafer, and W. Kunz (1996). Direct PCR on fruitflies and blood flukes without prior DNA isolation. *Nucleic Acids Res.* **24**: 4100–4101.

Gutierrez, G., and A. Marin (1998). The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Mol. Biol. Evol.* **15**: 926–929.

Guyer, R. L., and D. E. Koshland, Jr. (1989). The molecule of the year. *Science* **246**: 1543–1546.

Hadrys, H., M. Balick, and B. Schierwater (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.* **1**: 55–63.

Hadrys, H., B. Schierwater, S. L. Dellaporta, R. DeSalle, and L. W. Buss (1993). Determination of paternity in dragonflies by random amplified polymorphic DNA fingerprinting. *Mol. Ecol.* **2**: 79–87.

Hallden, C., M. Hansen, N. O. Nilsson, and A. Hjerdin (1996). Competition as a source of errors in RAPD analysis. *Theor. Appl. Genet.* **93**: 1185–1192.

Hammond, J. B. W., G. Spanswick, and J. A. Mawn (1996). Extraction of DNA from preserved animal specimens for use in randomly amplified polymorphic DNA analysis. *Anal. Biochem.* **240**: 298–300.

Handt, O., M. Hoss, M. Krings, and S. Paabo (1994). Ancient DNA: methodological challenges. *Experientia* **50**: 524–529.

Harris, S., and D. B. Jones (1997). Optimization of the polymerase chain reaction. *Brit. J. Biomed. Sci.* **54**: 166–173.

He, Q., M. K. Viljanen, and J. Mertsola (1994a). Effects of thermocyclers and primers on the reproducibility of banding patterns in randomly amplified polymorphic DNA analysis. *Mol. Cell. Probes* **8**: 155–160.

He, Q., M. Marjamaki, H. Soini, J. Mertsola, and M. K. Viljanen (1994b). Primers are decisive for sensitivity of PCR. *BioTechniques* **17**: 82–86.

Hedges, S. B., and M. H. Schweitzer (1995). Detecting dinosaur DNA. *Science* **268**: 1191.

Her, C., and R. M. Weinshilboum (1995). Rapid restriction mapping by use of Long PCR. *BioTechniques* **19**: 530–532.

Higgins, J. A., and A. F. Azad (1995). Use of polymerase chain reaction to detect bacteria in arthropods: a review. *J. Med. Entomol.* **32**: 213–222.

Hinnebusch, B. J., K. L. Gage, and T. G. Schwan (1998). Estimation of vector infectivity rates for plague by means of a standard curve-based competitive polymerase chain reaction method to quantify *Yersinia pestis* in fleas. *Am. J. Trop. Med. Hyg.* **58**: 562–569.

Hofreiter, M., D. Serre, H. N. Poinar, M. Kuch, and S. Paabo (2001). Ancient DNA. *Nat. Rev. Genet.* **2**: 353–359.

Hoy, M. A., A. Jeyaprakash, R. Morakote, P. K. C. Lo, and R. Nguyen (2000). Genomic analyses of two populations of *Ageniaspis citricola* (Hymenoptera: Encyrtidae) suggest that a cryptic species may exist. *Biological Control* **17**: 1–10.

Hoy, M. A., A. Jeyaprakash,, and R. Nguyen (2001). Long PCR is a sensitive method for detecting *Liberobacter asiaticum* in parasitoids undergoing risk assessment in quarantine. *Biol. Control* **22**: 278–287.

Iizuka, M., Y. Sugiyama, S. Iida, and T. Seikiya (1996). Direct sequencing of DNA fragments amplified with single primers by polymerase chain reaction without cloning. *Anal. Biochem.* **241**: 136–139.

Innis, M. A., D. H. Gelfand, J. J. Ninsky, and T. J. White, Eds. (1990). *PCR Protocols, A Guide to Methods and Applications*. Academic Press, San Diego.

Jackson, D. P., J. D. Hayden, and P. Quirke (1991). Extraction of nucleic acid from fresh and archival material, pp. 29–50. In: *PCR. A Practical Approach*. M. J. McPherson, P. Quirke, and G. R. Taylor, Eds. IRL, Oxford.

Jeyaprakash, A., and M. A. Hoy (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* **9**: 393–405.

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer, and B. Vosman, et al. (1997). Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breeding* **3**: 381–390.

Jowett, T. (1986). Preparation of nucleic acids, pp. 275–286. In: *Drosophila*, *A Practical Approach*. D. B. Roberts, Ed., IRL, Oxford.

Kainz, P. (2000). The PCR plateau phase—toward an understanding of its limitations. *Biochim. Biophys.* **1494**: 23–27.

Kalinina, O., I. Lebedeva, J. Brown, and J. Silver (1997). Nanoliter scale PCR with *Taq*Man detection. *Nucleic Acids Res.* **25**: 1999–2004.

Kambhampati, S., and P. T. Smith (1995). PCR primers for the amplification of four insect mitochondrial gene fragments. *Insect Mol. Biol.* **4**: 233–236.

Kapsa, R., N. Siregar, A. Quigley, J. Ojaimi, S. Katsabanis, C. Sue, and E. Byrne (1997). The polymerase chain reaction in the study of mitochondrial genetics. *J. Biochem. Biophys. Methods* **36**: 31–50.

Karl, S. A., and J. C. Avise (1993). PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Mol. Biol. Evol.* **10**: 342–361.

Kaufman, D. L., and G. A. Evans (1990). Restriction cleavage at the termini of PCR products. *BioTechniques* **8**: 304–306.

Kawasaki, E. S. (1990). Amplification of RNA, pp. 21–28. In: *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York.

Keohavong, P., and W. G. Thilly (1989). Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. USA* **86**: 9253–9257.

Kernodle, S. P., R. E. Cannon, and J. G. Scandalios (1993). Concentration of primer and template qualitatively affects products in Random-Amplified Polymorphic DNA PCR. *BioTechniques* **14**: 362–364.

Khandka, D. K., M. Tuna, M. Tal, A. Nejidat, and A. Golan-Goldhirsh (1997). Variability in the pattern of random amplified polymorphic DNA. *Electrophoresis* **18**: 2852–2856.

Kitchin, P. A. Z. Szotyori, C. Fromholc, and N. Almond (1990). Avoidance of false positives. *Nature* **344**: 201.

Knipple, D. C., L. L. Payne, and D. M. Soderlund (1991). PCR-generated conspecific sodium channel gene probe for the house fly. Arch. *Insect Bioch. Physiol.* **16**: 45–53.

Kocher, T. D. (1992). PCR, direct sequencing, and the comparative approach. *PCR Methods Appl.* **1**: 217–221.

Kubota, T., S. Das, S. L. Christian, S. B. Baylin, J. G. Herman and D. H. Ledbetter (1997). Methylation-specific PCR simplifies imprinting analysis. *Nat. Genet.* **16**: 16–17.

Lamboy, W. F. (1994). Computing genetic similarity coefficients from RAPD data: the effects of PCR artifacts. *PCR Methods Appl.* **4**: 31–37.

Lanciotti, R. S., J. T. Roehrig, V. Deubel, J. Smith, M. Parker, K. Steele, B. Crise, K. E. Volpe et al. (1999). Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. *Science* **286**: 2333–2337.

Lanciotti, R. S., A. J. Kerst, R. S. Nasci, M. S. Dogsey, C. J. Mitchell, H. M. Savage, N. Komar, N. A. Panella, B. C. Allen, K. E. Volpe, B. S. Davis, and J. T. Roehrig (2000). Rapid detection of West Nile virus from human clinical specimens, field-collected mosquitoes and avian samples by a TaqMan reverse transcriptase-PCR assay. *J. Clin. Microbiol.* **38**: 4066–4071.

Landweber, L. F., and M. Kreitman (1995). Producing single-stranded DNA in polymerase chain reaction for direct genomic sequencing, pp. 579–588. In: *Recombinant DNA Methodology II*. Academic Press, San Diego.

Laurent, V., E. Wajnberg, B. Mangin, T. Schiex, C. Gaspin, and F. Vanlerberghe-Masutti (1998). A composite genetic map of the parasitoid wasp *Trichogramma brassicae* based on RAPD markers. *Genetics* **150**: 275–282.

Lee, A. B., and T. A. Cooper (1995). Improved direct PCR screen for bacterial colonies: wooden toothpicks inhibit PCR amplification. *BioTechniques* **18**: 225–226.

Lexa, M., J. Horak, and B. Brzobohaty (2001). Virtual PCR. *Bioinformat. Appl. Note* **17**: 192–193.

Lie, X., and R. P. Singh (2001). A novel usage of random primers for multiplex RT-PCR detection of virus and viroid in aphids, leaves and tubers. *J. Virol. Methods* **91**: 37–49.

MacPherson, J. M., P. E. Eckstein, G. J. Scoles, and A. A. Gajadhar (1993). Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Mol. Cell. Probes* **7**: 293–299.

Massung, R. F., K. Slater, J. H. Owens, W. L. Nicholson, T. N. Mather, V. B. Solberg, and J. G. Olson (1998). Nested PCR assay for detection of granulocytic Ehrlichiae. *J. Clin. Microbiol.* **36**: 1090–1095.

Mauel, M. J., S. J. Carlton, and T. N. Mather (1999). Polymerase chain reaction detection efficiency of the human granulocytic ehrlichiosis agent (Rickettsiaceae: Ehrlichieae) in ticks (Acari: Ixodidae) is dependent on the DNA extraction method. *J. Med. Entomol*. **36**: 649–652.

McClelland, M., and J. Welsh (1994). DNA fingerprinting by arbitrarily primed PCR. *PCR Methods Appl.* **4**: 559–565.

McClelland, M., F. Mathieu-Daude, and J. Welsh (1995). RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends Genet.* **11**: 242–245.

McEwan, N. R., J. Bakht, E. Cecchini, A. Coultart, C. Geri, F. M. McDonald, M. S. McDonald, N. N. Ab Rahman, and G. Williamson (1998). Examination of the pipetting variation between individuals performing random amplified polymorphic DNA (RAPD) analysis. *J. Microbiol. Methods* **32**: 213–215.

McPherson, M. J., P. Quirke, and G. R. Taylor, Eds. (1991). *PCR. A Practical Approach*. IRL, Oxford.

McPherson, M. J., K. M. Jones, and S. J. Gurr (1991). PCR with highly degenerate primers, pp. 171–186. In: *PCR. A Practical Approach*. M. J. McPherson, P. Quirke, and G. R. Taylor, Eds. IRL, Oxford.

Mertz, L. M., and A. Rashtchian (1994). Nucleotide imbalance and polymerase chain reaction: effects on DNA amplification and synthesis of high specific activity radiolabeled DNA probes. *Anal. Biochem.* **221**: 160–165.

Meunier, J. R., and P. A. D. Grimont (1993). Factors affecting reproducibility of random amplified polymorphic DNA fingerprinting. *Res. Microbiol.* **144**: 373–379.

Micheli, M. R. R. Bova, P. Calissano, and E. D'Ambrosio (1993). Randomly amplified polymorphic DNA fingerprinting using combinations of oligonucleotide primers. *BioTechniques* **15**: 388–390.

Micheli, M. R., R. Bova, E. Pascale, and E. D'Ambrosio (1994). Reproducible DNA fingerprinting with the random amplified polymorphic DNA (RAPD) method. *Nucleic Acids Res.* **22**: 1921–1922.

Moreira, D. (1998). Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. *Nucleic Acids Res.* **26**: 3309–3310.

Morlais, I., P. Grebaut, J. M. Bodo, S. Djoha, and G. Cuny (1998). Characterization of trypanosome infections by polymerase chain reaction (PCR) amplification in wild tsetse flies in Cameroon. *Parasitology* **116**: 547–554.

Mueller, U. G., and L. L. Wolfenbarger (1999). AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* **14**: 389–394.

Mullis, K. B. (1987). Process for amplifying nucleic acid sequences. U.S. Patent 4,683,202. Filed October 25, 1985; issued July 27, 1987 .

Mullis, K. B., and F. Faloona (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155**: 335.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor. *Symp. Quant. Biol.* **51**: 263.

Nelson, W. S., P. A. Prodohl, and J. C. Avise (1996). Development and application of long-PCR for the assay of full-length animal mitochondrial DNA. *Mol. Ecol.* **5**: 807–810.

Nevett, C. L., and A. Louwrier (2000). A formulation of multiple reverse transcriptases for improved RT-PCR. *Biotechnol. Lett.* **22**: 1774–1778.

Nie, X., and R. P. Singh (2001). A novel usage of random primers for multiplex RT-PCR detection of virus and viroid in aphids, leaves and tubers. *J. Virol. Methods* **91**: 37-49.

Ochman, H., M. M. Medhora, D. Garza, and D. L. Hartl (1990). Amplification of flanking sequences by inverse PCR, pp. 219–227. In: *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York.

Olsen, D. B., G. Wunderlich, A. Uy, and F. Eckstein (1993). Direct sequencing of polymerase chain reaction products, pp. 79–92. In: *Methods in Enzymology.* Vol. 218. *Recombinant DNA. Part I*. R. Wu, Ed., Academic Press, San Diego.

Osborne, B. I. (1992). HyperPCR: a Macintosh HyperCard program for the determination of optimal PCR annealing temperature. *CABIOS* **8**: 83. Oxford Univ. Press.

Paabo, S. (1990). Amplifying ancient DNA, pp. 159–166. In: *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York.

Paabo, S. (1991). Amplifying DNA from archeological remains: a meeting report. *PCR Methods Appl.* **1**: 107–110.

Pao, C. C., J. J. Hor, P. L. Tsai, and M. Y. Horng (1993). Inhibition of *in vitro* enzymatic DNA amplification reaction by ultra-violet light irradiation. *Mol. Cell. Probes* **7**: 217–239.

Perez, T., J. Albornoz, and A. Dominguez (1998). An evaluation of RAPD fragment reproducibility and nature. *Mol. Ecol.* **7**: 1347–1357.

Perler, F. B., S. Kumar, and H. Kong (1996). Thermostable DNA polymerases. *Adv. Protein Chem.* **48**: 377–435.

Persing, D. H., S. R. Telford III, P. N. Rys, D. E. Dodge, T. J. White, S. E. Malawista, and A. Spielman (1990). Detection of *Borrelia burgdorferi* DNA in museum specimens of *Ixodes dammini* ticks. *Science* **249**: 1420–1423.

Peter, T. F., A. F. Barbet, A. R. Alleman, B. H. Simbi, M. J. Burridge, and S. M. Mahan (2000). Detection of the agent of heartwater, *Cowdria ruminantium*, in *Amblyomma* ticks by PCR: validation and application of the assay to field ticks. J. *Clin. Microbiol.* **38**: 1535–1544.

Poinar, H. N., and B. A. Stankiewicz (1999). Protein preservation and DNA retrieval from ancient tissues. *Proc. Natl. Acad. Sci. USA* **96**: 8426–8431.

Priest, F. G. (1995). Age of bacteria from amber. *Science* **270**: (2015).

Quicke, D. L. J., R. Belshaw, and C. Lopez-Vaamonde (1999). Preservation of hymenopteran specimens for subsequent molecular and morphological study. *Norweg. Acad. Sci. Lett. Zool. Scripta* **28**: 261–267.

Rabinow, P. (1996). *Making PCR. A Story of Biotechnology*. Univ. of Chicago Press.

Rabinow, L., A. Barrera, and J. Hyde (1993). A simple method for isolation of very high molecular weight DNA from *Drosophila* embryos. *Nucleic Acids Res.* **21**: 4985–4986.

Rameckers, J., S. Hummel, and B. Herrmann (1997). How many cycles does a PCR need? Determinations of cycle numbers depending on the number of targets and the reaction efficiency factor. *Naturwissenschaften* **84**: 259–262.

Rao, V. B. (1994a). Direct sequencing of polymerase chain reaction-amplified DNA. *Anal. Biochem.* **216**: 1–14.

Rao, V. B. (1994b). Strategies for direct sequencing of PCR-amplified DNA. *PCR Methods Appl.* **4**: 515–523.

Reineke, A., P. Kaarlovsky, and C. P. W. Zebitz (1998). Preparation and purification of DNA from insects for AFLP analysis. *Insect Mol. Biol.* **7**: 95–99.

Roehrdanz, R. L. (1995). Amplification of complete insect mitochondrial genome in two easy pieces. *Insect Mol. Biol.* **4**: 169–172.

Rollo, F. (1998). Ancient DNA: problems and perspectives for molecular microbial paleoecology, pp. 133–149. In: *Advances in Molecular Ecology*. G. R. Carvalho, Ed. IOS Press, Amsterdam.

Roux, K. H. (1995). Optimization and troubleshooting in PCR. *PCR Methods Appl.* **4**: S185–S194.

Ruano, G., D. E. Brash, and K. K. Kidd (1991). PCR: The first few cycles. Amplifications. A Forum for PCR Users. PerkinElmer Cetus, Norwalk, CT.

Rychlik, W. W. J. Spencer, and R. E. Rhoads (1990). Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res.* **18**: 6409–6412.

Saiki, R. K., S. J. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.

Sall, T., C. Lind-Hallden, and C. Hallden (2000). Primer mixtures in RAPD analysis. *Hereditas* **132**: 203–208.

Sambrook, J., and D. W. Russell (2001). *Molecular Cloning. A Laboratory Manual*. Vol. 2, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Savelkoul, P. H. M., H. J. M. Aarts, J. De Haas, L. Dijkshoorn, B. Duim, M. Otsen, J. L. W. Rademaker, L. Schouls, and J. A. Lenstra (1999). Amplified-fragment length polymorphism analysis: the state of the art. *J. Clin. Microbiol.* **37**: 3083–3091.

Scharf, S. J. (1990). Cloning with PCR, pp. 84–91. In: *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York.

Schierwater, B. (1995). Arbitrarily amplified DNA in systematics and phylogenetics. *Electrophoresis* **15**: 1643–1647.

Schierwater, B., and A. Ender (1993). Different thermostable DNA polymerases may amplify different RAPD products. *Nucleic Acids Res.* **21**: 4647–4648.

Schriefer, M. E., J. B. Sacci, Jr., R. A. Wirtz, and A. F. Azad (1991). Detection of polymerase chain reaction-amplified malarial DNA in infected blood and individual mosquitoes. *Exp. Parasitol.* **73**: 311–316.

Schwartz, I., S. Varde, R. B. Nadelman, G. P. Wormser, and D. Fish (1997). Inhibition of efficient polymerase chain reaction amplification of *Borrelia burgdorferi* DNA in blood-fed ticks. *Am. J. Trop. Med. Hyg.* **56**: 339–342.

Schweder, M. E., R. G. Shatters, Jr., S. H. West, and R. L. Smith (1995). Effect of transition interval between melting and annealing temperatures on RAPD analyses. *BioTechniques* **19**: 38–42.

Shahjahan, R. M., K. J. Hughes, R. A. Leopold, and J. D. DeVault (1995). Lower incubation temperature increases yield of insect genomic DNA isolated by the CTAB method. *BioTechniques* **19**: 333–334.

Siebert, P. D., and J. W. Larrick (1992). Competitive PCR. *Nature* **359**: 557–558.

Silver, J. (1991). Inverse polymerase chain reaction, pp. 137–146. In: *PCR. A Practical Approach*, M. J. McPherson, P. Quirke and G. R. Taylor, Eds. IRL, Oxford.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook (1994). Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**: 651–701.

Singer-Sam, J., R. L. Tanguay, and A. D. Riggs (1989). Use of Chelex to improve the PCR signal from a small number of cells. Amplifications 1989, no. 3, p. 11.

Steiner, J. J., C. J. Poklemba, R. G. Fjellstrom, and L. F. Elliott (1995). A rapid one-tube genomic DNA extraction process for PCR and RAPD analyses. *Nucleic Acids Res.* **23**: 2569–2570.

Suazo, A., and H. G. Hall (1999). Modification of the AFLP protocol applied to honey bee (*Apis mellifera* L.) DNA. *BioTechniques* **26**: 704–708.

Sykes, B. (1997). Lights turning red on amber. *Nature* **386**: 764–765.

Taylor, G. R. (1991). Polymerase chain reaction: basic principles and automation, pp. 1–14 . In: *PCR. A Practical Approach*, M. J. McPherson, P. Quirke, and G. R. Taylor, Eds., IRL, Oxford.

Tingey, S. V., and J. P. del Tufo (1993). Genetic analysis with random amplified polymorphic DNA markers. *Plant Physiol.* **101**: 349–352.

Townson, H., R. E. Harbach, and T. A. Callan (1999). DNA identification of museum specimens of the *Anopheles gambiae* complex: an evaluation of PCR as a tool for resolving the formal taxonomy of sibling species complexes. *Syst. Entomol.* **24**: 95–100.

Vanlerberghe-Masutti. F. (1994). Molecular identification and phylogeny of parasitic wasp species (Hymenoptera: Trichogrammatidae) by mitochondrial DNA RFLP and RAPD markers. *Insect Mol. Biol.* **3**: 229–237.

Vogler, A. P., and D. L. Pearson (1996). A molecular phylogeny of the tiger beetles (Cicindelidae): congruence of mitochondrial and nuclear rDNA data sets. *Mol. Phyl. Evol.* **6**: 321–338.

Vos, P. R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**: 4407–4414.

Wada, M., C. Klein, J. Schell, and B. Reiss (1994). A functional assay for *Taq* DNA polymerase in PCR. *BioTechniques* **16**: 28–30.

Walden, K. K. O., and H. M. Robertson (1997). Ancient DNA from amber fossil bees? *Mol. Biol. Evol.* **14**: 1075–1077.

Waters, L. C., S. C. Jacobson, N. Kroutchinina, J. Khandurina, R. S. Foote, and J. M. Ramsey (1998). Multiple sample PCR amplification and electrophoretic analysis on a microchip. *Anal. Chem.* **70**: 5172–5176.

Weill, M., C. Berticat, M. Raymond, and C. Chevillon (2000). Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. *Anal. Biochem.* **285**: 267–270.

Welsh, J., and M. McClelland (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18**: 7213–7218.

Welsh, J., C. Petersen, and M. McClelland (1990). Polymorphisms generated by arbitrarily primed PCR in the mouse; application to strain identification and genetic mapping. *Nucleic Acids Res*. **19**: 303–306.

Welsh, J., K. Chada, S. S. Dalal, R. Cheng, D. Ralph, and M. McClelland (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res.* **20**: 4965–4970.

White, T. J., N. Arnheim, and H. A. Erlich (1989). The polymerase chain reaction. *Trends Genet*. **5**: 185–189.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18**: 6531–6535.

Williams, J. G. K., M. K. Hanafey, J. A. Rafalski, and S. V. Tingey (1993). Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.* **218**: 704–740.

Wise, D. L., and T. L. Weaver (1991). Detection of the Lyme disease bacterium, *Borrelia burgdorferi*, by using the polymerase chain reaction and a nonradioisotopic gene probe. *J. Clin. Microbiol.* **29**: 1523–1526.

Wright, D. K., and M. M. Manos (1990). Sample preparation from paraffin-embedded tissues, pp. 153–158. In: *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York.

Yousten, A. A., and K. E. Rippere (1997). DNA similarity analysis of a putative ancient bacterial isolate obtained from amber. FEMS *Microbiol. Lett.* **152**: 345–347.

9

Transposable-Element Vectors to Transform *Drosophila* and Other Insects

9.1. Overview

P transposable elements (TEs) have been genetically modified to serve as vectors for inserting exogenous DNA into *Drosophila*. This new tool has revolutionized research on insect development and analyses of gene structure, function, regulation, and position effects in *D. melanogaster*.

P elements are found in certain strains (called P) of *D. melanogaster* but are lacking in others (called M). When P males and M females are crossed, the F_1 progenies exhibit a condition called "hybrid dysgenesis" because the *P* elements present in the chromosomes of the F_1 progenies are no longer prevented from transposing or moving. The resultant insertions of *P* elements into new chromosomal sites lead to mutations and sterility (= hybrid dysgenesis). *P* elements appear to have invaded *D. melanogaster* about 50 years ago from another *Drosophila* species (an example of horizontal transmission). One possible method by which *P* could have been transferred into *D. melanogaster* is by a mite. It was hypothesized that *P* elements were obtained from the eggs of one *Drosophila* species by a parasitic mite during feeding. Subsequent feeding by this "infected" mite on *D. melanogaster* eggs might have resulted in the mechanical transfer of *P* elements to *D. melanogaster* and their subsequent spread in field populations around the world. Alternatively, rare interspecies matings could have allowed *P* to invade *D. melanogaster*.

When *P*-element vectors containing cloned genes are microinjected into early-stage *Drosophila* embryos, some of the *P* vectors integrate into the chromosomes in germ-line tissues. If the newly inserted DNA is transmitted to the progeny of the injected embryos, stable transformation has occurred.

Transposon tagging, which occurs when a single *P* inserts into a gene and causes a visible mutation, facilitates the identification and cloning of genes from *Drosophila*. A gene is "tagged" with the *P*, making it easier to clone it. *P*-element vectors repair gaps left in chromatids when *P* elements excise, which offers the possibility of inserting exogenous DNA into targeted, rather than random, sites in the *Drosophila* chromosome.

Many drosophilid species have inactive forms of *P* in their genomes. Inactive elements are suppressed by several mechanisms to reduce the deleterious effects active transposition imposes on individuals and populations. Thus, the long-term survival of TEs such as *P* may require that they move horizontally into new species.

Although *P*-element vectors have not been useful for transforming insects other than *Drosophila*, other TEs, including *Hermes*, *minos*, *hobo*, *piggyBac* and *mariner*, have been engineered as vectors and used to transform a variety of insect species, thus providing tools with which to insert exogenous genes into both pest and beneficial insects with the goal of improving pest management programs.

9.2. Introduction

The *P* element first was genetically modified to serve as a vector of exogenous DNA in 1982 (Spradling and Rubin 1982, Rubin and Spradling 1982). A variety of different

P-element vectors now are used routinely to introduce exogenous DNA into *Drosophila melanogaster*. *P*-mediated transformation of *D. melanogaster* has revolutionized how geneticists study gene structure, function, regulation, position effects, dosage compensation, and development. *P*-mediated transformation allows geneticists to decipher the genetic basis of behavior, development, and sex determination in *Drosophila*, as will be described in Chapters 10 and 11.

This chapter describes *P* elements and hybrid dysgenesis, and the methods employed in introducing *P*-element vectors into the germ line of *D. melanogaster*. This approach to germ-line transformation has inspired entomologists to attempt to engineer insects other than *Drosophila* using TE vectors with a broader host range.

An analysis of the spread of *P* elements into natural populations of *D. melanogaster* previously lacking these TEs allows us to learn more about the evolution of *P* and other TEs. Because TEs have been proposed as possible "drive mechanisms" for the genetic engineering of wild insect populations, the study of *P*-element invasion into *D. melanogaster* may serve as a model to understand the potential for using this type of drive mechanism for other insects. Analyses of *P*-element invasion also may provide clues to the evolution of resistance (suppressive factors) to TEs.

9.3. *P* Elements and Hybrid Dysgenesis

Intact *P* elements are 2907 bp long and encode a single polypeptide that has transposase activity (Figure 9.1). There are four exons (numbered from 0 to 3) flanked by inverted repeats 31 bp long. The presence of intact inverted repeats is required if the *P* element is to transpose (move).

Multiple copies of *P* (30 to 60) are dispersed throughout the genome of P strains of *D. melanogaster*, but are not active because transposition is suppressed by factors in the P cytotype. Many *P* elements in *D. melanogaster*, and other *Drosophila* species, have some sequences deleted, which also makes them incapable of transposing.

Movements of *P* elements cause mutations by inactivating genes, altering rates of transcription or development- or tissue-specific gene expression. *P*-element movements break chromosomes and cause nondisjunction during meiosis that can lead to chromosome rearrangements and germ-cell death. Transposition of *P* elements in somatic cells reduces

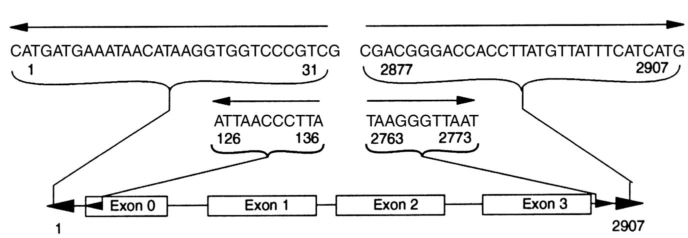


Figure 9.1. Structure of an intact *P* element. There are four exons (0–3), separated by short introns (thin line). The 31-bp inverted terminal repeats (sequences 1–31 and 2877–2907) are indicated by the filled arrows. There are also inverted repeats at sequences 126–136 and 2763–2773 (modified from Engels 1989).

the life span of *D. melanogaster* males, as well as reducing fitness, mating activity, and locomotion (Woodruff et al. 1999).

A syndrome called **hybrid dysgenesis** is induced in *D. melanogaster* when males from a strain that contains *P* (P males) are mated with females lacking *P* (M females) (Bingham et al. 1982). Their F_1 progenies exhibit a high rate of mutation, chromosomal aberrations, and, sometimes, complete sterility, caused by transposition of *P* in their germ-line chromosomes. The reciprocal cross does not exhibit these negative effects because the P female's cytotype suppresses movement of *P*.

When transposition of *P* elements occurs in germ-line tissues in *D. melanogaster*, three short introns have to be cut out of the original RNA transcript to produce the mRNA that codes for a functional 87-kilodalton (kDa) transposase (Laski et al. 1986, Rio et al. 1986, Kobayashi et al. 1993). In somatic cells, a 66-kDa protein is produced that can function as a repressor of *P* activity (Lemaitre et al. 1993).

9.4. *P*-Element Structure Varies

Many *P* elements in the *Drosophila* genome are defective. Some have internal deletions and are unable to produce their own transposase but, if they retain their 31-bp terminal repeats, they can move if supplied with transposase by intact elements. An early report by Eggleston et al. (1988) suggested that *P* elements are unable to mobilize other TE families in *D. melanogaster*. However, that may not be true, because *P* elements could mobilize the *Tc1* element from the nematode *C. elegans* (Szekely et al. 1994).

In fact, diverse TEs can be mobilized in dysgenic crosses between strains of *D. virilis* (Petrov et al. 1995). Four different TEs, *Ulysses*, *Penelope*, *Paris* and *Helena*, were mobilized in the dysgenic crosses despite the fact that these TEs are structurally diverse. *Ulysses* is a retroelement related to the *Ty3-gypsy* superfamily, and *Penelope* is similar to another class of retroelements. *Paris* is in the *mariner/Tc1* superfamily that transposes without an RNA intermediate, while *Helena* is a LINE-like element. There was no evidence that the simultaneous mobilization was due to complementation of some shared defect in the transposition pathways; the transposition mechanisms employed by these elements are different, and the mutants examined all showed evidence that transposition occurred in the appropriate manner. The four TEs appear to have been mobilized due to the "genomic stress" brought about by the dysgenic cross (Petrov et al. 1995). The stress could have been caused by breakage of double-stranded DNA. Double-stranded DNA breakage is caused by exposure to UV light and other agents and can increase transcription and/or mobilization of some retroelements. Thus, the production of ds breaks from the mobilization of a single TE might induce a cellular response that releases other TEs from repression, allowing a single system of hybrid dysgenesis to mobilize multiple, unrelated elements.

P elements with defective 31-bp terminal repeats are unable to transpose because these repeats are the site of action of the transposase. The frequency of transposition depends on the size of the *P*; smaller elements are able to move more readily than larger ones. The location of the *P* in the chromosome also is important in determining the frequency of transposition. Although transposition is more or less "random" at the genome scale, *P* elements containing specific gene sequences show some specificity by frequently inserting near the parent gene (which is called "homing") (Taillebourg and Dura 1999).

P elements also tend to insert into upstream promoter regions of genes (Spradling et al. 1995).

9.5. Transposition Method of *P* Elements

P elements move from site to site in the genome (jump) by a "cut and paste" method (Engels et al. 1990, Gloor et al. 1991, Sentry and Kaiser 1992, Engels 1997). When a *P* jumps, it leaves behind a ds gap in the DNA. The gap is repaired by using a matching sequence as a template. This matching sequence can occur on the sister chromatid or elsewhere in the genome. If the transposition occurs in an individual that is *heterozygous* for the *P* insertion, and the matching site on the homologous chromosome is used as the template for DNA replication and repair, there can be a precise *loss* of the *P* sequence in the original site, although there is no *net loss* in the genome because the *P* element has simply changed locations.

However, if a *P* jumps after the chromosomes have duplicated, but before the cell divides, one of the sister chromatids will still have a *P* in its original position. In this situation, this homologous *P* may serve as the template for filling in the hole left when the *P* moved to a new position in the genome. Under these circumstances, the number of *P* elements in the genome is *increased* by one. The *P* element is *replaced* in its original site by gap repair *and* also is present in a new site in the genome.

The "cut and paste" mechanism of transposition implies that *P* elements don't have to confer an advantage on the organism to invade and persist in the genome. In fact, a mathematical simulation model indicates that *P* elements can become fixed in populations even when fitness is reduced by 50% (Hickey 1982) and many laboratory studies have shown that colonies can change rapidly from M to P strains. The cut and paste model has been the conceptual basis of **targeted gene replacement** in *Drosophila* (Engels et al. 1990, Gloor et al. 1991, Sentry and Kaiser 1992, Rong and Golic 2000), which will be described.

9.6. Origin of *P* Elements in *D. melanogaster*

P elements are relatively new to *D. melanogaster*. Surveys indicate laboratory strains of *D. melanogaster* collected before 1950 lack *P*, but most colonies collected from the wild within approximately the past 50 years have *P* elements (Anxolabehere et al. 1988, Engels 1989, 1992, Powell and Gleason 1996).

P elements are relatively common in other species of *Drosophila*. Surveys indicated that closely related, full-sized, and potentially active *P* elements occur in *D. willistoni*, *D. guanche*, *D. bifasciata*, and *Scaptomyza pallida* (Hagemann et al. 1996). A *P* element isolated from *Scaptomyza pallida*, a drosophilid distantly related to *D. melanogaster*, is able to transpose in *D. melanogaster* and to mobilize a defective *D. melanogaster P* element (Simonelig and Anxolabehere 1991).

Phylogenetic analyses of DNA sequences from *P* elements in 17 *Drosophila* species in the *melanogaster* species group within the subgenus *Sophophora* show that sequences from the *P* family fall into distinct subfamilies or clades which are characteristic for particular species subgroups (Clark and Kidwell 1997, Clark et al. 1998). These clades indicate that vertical transmission of *P* elements has occurred, but in some cases the *P* phylogeny is *not* congruent with species phylogeny. More than one subfamily of *P* elements may exist

within a group, with sequences differing by as much as 36%, suggesting that horizontal transfer has occurred. In fact, horizontal transfer may be essential to the long-term survival of TEs (Clark and Kidwell 1997).

P probably invaded *D. melanogaster* within the past 50 years. The donor species that provided a *P* to *D. melanogaster* is thought to be in the *willistoni* group, which is not closely related to *D. melanogaster* (Daniels et al. 1984, Lansman et al. 1985, Daniels and Strausbaugh 1986, Engels 1997). Because these species diverged from each other about 60 million years ago, there should have been sufficient time for considerable sequence divergence in the *P* elements if they had been present in both genomes prior to divergence (and subsequently transmitted vertically). However, *P*-element sequences from *melanogaster* and *willistoni* are nearly identical, supporting the hypothesis of horizontal transfer. Engels (1997) speculated that the invasion of *D. melanogaster* by *P* occurred after *D. melanogaster* was introduced into the Americas, and that invasion by TEs could be a "general hazard associated with the expansion of any species into a new ecosystem." Such TE invasions potentially could provide genetic variation that contributes to postrelease adaptations that occur in some species subsequent to their invasion into new environments.

Two mechanisms have been proposed to explain how *P* could have infected *D. melanogaster*. One involves horizontal transfer, and the other involves interspecific crosses. Both *D. melanogaster* and *D. willistoni* now overlap in their geographical ranges in Florida and in Central and South America, but they apparently are unable to interbreed. Horizontal transfer could have been effected by a viral, bacterial, fungal, protozoan, spiroplasmal, mycoplasmal, or small arthropod vector (Hymenoptera or Acari (mites)). One candidate for horizontal vector may be a semiparasitic mite, *Proctolaelaps regalis* (Houck et al. 1991, Kidwell 1992, Engels 1997). *P. regalis* is associated with both *Drosophila* species; it has been found in laboratory colonies and in the field associated with fallen or rotting fruit, which is the natural habitat for *Drosophila*. Laboratory observations indicate that *P. regalis* feeds on fly eggs, larvae, and pupae and can make rapid thrusts of its mouth parts into a series of adjacent hosts. This brief feeding on multiple hosts might allow it to pick up DNA from one egg and inject it into another. Mites from colonies of *Drosophila* with *P* elements in their genome were analyzed by the PCR and Southern blot analysis. The analyses indicated the mites carried both *P* and *Drosophila* ribosomal DNA sequences. Mites isolated from M colonies (which lack *P*) lacked *P* sequences.

For *P. regalis* to have transferred *P* elements from *D. willistoni* to *D. melanogaster*, a number of conditions had to occur in the proper sequence (Houck 1993). Females of *D. melanogaster* and *D. willistoni* had to deposit their eggs in close proximity, and mites had to feed sequentially on one and then the other, in the correct order. The recipient egg had to be less than 3 hours old, the germ line of the recipient embryo had to incorporate a complete copy of the exogenous *P*, the transformed individual had to survive to adulthood, and the adult had to reproduce.

A second potential mechanism for horizontal transfer of *P* involves interspecific crosses. Crosses between the sibling species *D. simulans* and *D. mauritiana* produce sterile males but fertile females. When F_1 females are backcrossed to males of either species, a few fertile males are produced. To determine whether interspecific transmission of *P* might occur, the two species were crossed and the hybrid progeny were evaluated by *in situ* hybridization of larval salivary glands and Southern blot (Montchamp-Moreau et al. 1991). The results indicated that the *P* element is able to pass from one species to another when the postmating sterility barrier is incomplete. Hybridization, although rare, occurs between some *Drosophila* species.

P elements have been found in other Dipteran families, including Opomizydae and Trixoscelididae (Anxolabehere and Periquet 1987). Inactive *P* elements were found in the sheep blowfly *Lucilia cuprina* (Calliphoridae) and the housefly *Musca domestica* (Muscidae) (Lee et al. 1999). The *P* elements in *M. domestica* differed from those in *D. melanogaster* by having two introns in exon 2 (as does the *P* from *L. cuprina*). The lack of a functional exon 3 in the housefly *P* likely is the basis for the element's inactivity. The presence of *P* elements in families other than Drosophilidae suggests that *P* elements may be more widely distributed than currently thought (Lee et al. 1999). This could be confirmed by the use of PCR primers that reflect the diversity of known *P* sequences in other insects.

9.7. *P* Vectors and Germ-Line Transformation

9.7.1. Protocols

After *P* elements were cloned (Rubin et al. 1982), they were genetically engineered to serve as vectors to insert exogenous DNA into the germ line of *D. melanogaster* (Rubin and Spradling 1982, Spradling and Rubin 1982). A number of different vectors with different characteristics have been produced subsequently (Fujioka et al. 2000). The following example provides a model for the procedures involved in *P*-mediated transformation of *Drosophila* (Figure 9.2):

(A) Construct or choose an appropriate *P* vector containing the DNA and marker gene(s) of interest. In this example, the marker gene is a wild-type version (*white*$^+$) of the mutant gene for white eyes. The wild-type allele is dominant over *white* so if a single copy of *white*$^+$ is present, the fly will have normal red eye color. This vector is unable to insert into the chromosome because it can't produce its own transposase.

Select a **helper plasmid**, such as pπ25.7wc, that contains a complete DNA sequence coding for transposase. This vector is unable to insert into *Drosophila* chromosomes by the normal transposition method because it lacks 23 bp of one terminal repeat, hence the designation wc for "wings clipped."

(B) Microinject both the vector and helper plasmid into embryos (G_0) from an appropriate host strain with white eyes. Embryos should be in the preblastoderm stage, when the embryo is still a syncytium.

(C) Mate the injected G_0 individuals that survive to adulthood with males or females that are homozygous for *white*. If the wild-type gene was inserted into the chromosome, then the progeny will have red eyes because they will be heterozygous (w^+/w).

(D) Mate G_1 progeny with white-eyed (*white*) flies to produce the next generation (G_2) with wild-type eyes.

(E) Select individual G_2 lines with wild-type eyes.

(F) Identify possible transformants containing single insertions at unique sites (single-insert lines) and verify insert structure.

(G) Analyze the properties of the transformed lines, including level of expression of the inserted DNA and stability of the transformed line.

(H) Cross the most useful lines to balancer stocks to enable the lines to be maintained in a stable condition.

Insertion of *P*-element vectors' DNA into germ-line chromosomes is enhanced if preblastoderm embryos are microinjected. At that stage, the cleavage nuclei are in a

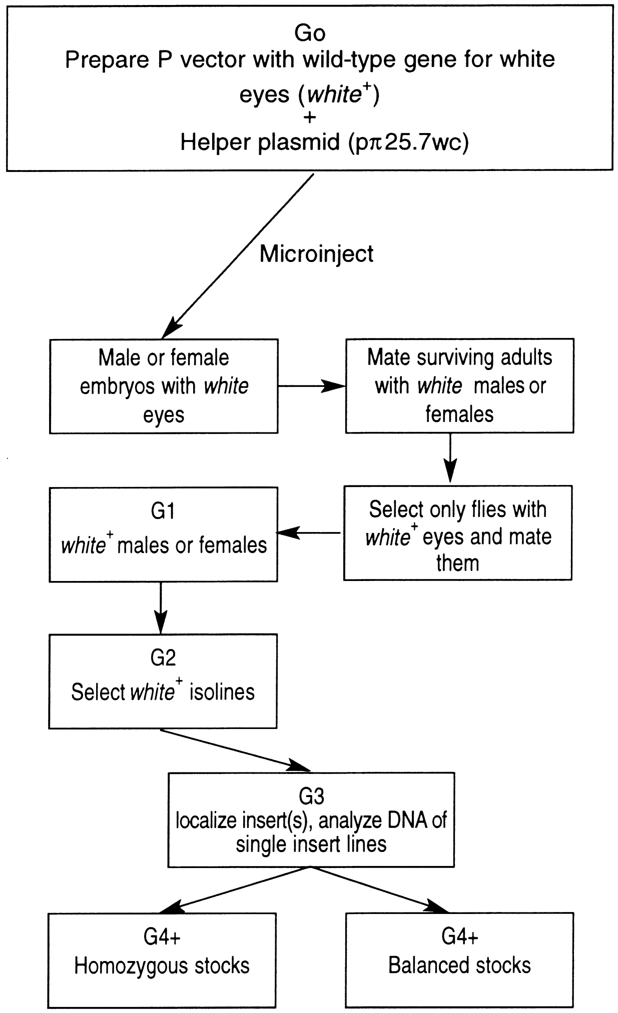


Figure 9.2. Steps in transforming *D. melanogaster* with *P*-element vectors. See text for details.

syncytium (lacking nuclear membranes) and the *P* elements can more easily be inserted into the chromosomes. The preblastoderm embryos are forming the **pole cells** that will give rise to the ovaries and testes. Insertion of exogenous DNA into the chromosomes of the germ line results in stable transformation. If only somatic cells contain the exogenous DNA, the flies cannot transmit the desired trait to their progeny. Such adult flies may exhibit the trait but are only transiently transformed.

Only a portion of the *P* vector inserts into the chromosome. The DNA inserted consists of the sequences contained within the inverted terminal repeats of the *P* element (Figure 9.3A). The plasmid DNA outside the inverted repeats should not insert and should be lost during subsequent development.

Once transformed fly lines are obtained, the lines typically are stable unless transposase is provided in some manner. The helper plasmid with transposase never inserts because it

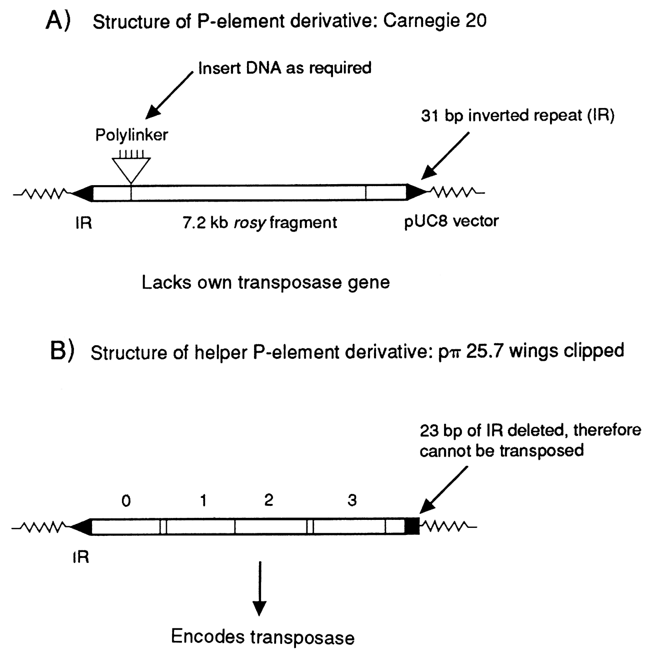


Figure 9.3. Examples of modified *P*-element vectors. (A) The Carnegie 20 vector contains a 7.2-kb segment of DNA coding for the *rosy* gene. It contains a polylinker for inserting exogenous DNA and retains the 31-bp inverted repeats (IR, dark arrows). This vector cannot transpose without a helper element because it cannot make transposase. (B) The helper element, pπ25.7 wings clipped, produces transposase, but 23 bp of inverted repeat has been deleted at one end so this vector cannot insert into the chromosome.

lacks normal inverted repeats (Figure 9.3B). Sometimes an experimenter wants to induce movements of inserted DNA, and **secondary transposition** can be induced if transposase is introduced by injecting helper elements into a preblastoderm embryo.

During embryogenesis and development of the injected larvae (which are the G_0 generation), transcription and translation of the wild-type gene can produce sufficient xanthine dehydrogenase to influence the eye color of the adults. (G_0 indicates the generation that is injected, G_1 indicates their progeny, and so on.) G_0 flies with a normal eye color do not necessarily have the *white*$^+$ gene inserted into the germ-line chromosomes because the injected DNA may be transcribed and translated while in the cytoplasm, or the DNA may only be inserted into the chromosomes of somatic cells (= **transient transformation**).

The next generation of flies (G_1) is the crucial generation to be screened for transformation, because these flies should have wild-type eyes *only* if the *white*$^+$ gene did insert into the germ-line chromosomes. The presence of one or more progeny with normal eye color in the G_1 indicates **stable transformation** occurred.

Individual G_1 flies may contain multiple insertions of the *P* element. Also, the *P* element may have inserted into different sites in different G_1 flies. As a result, colonies derived from single flies must be screened in order to identify colonies with a single insert (Figure 9.2).

To determine how many *P* elements inserted into the chromosomes of each colony and their location, DNA from G_2 adult flies is prepared from each isoline and evaluated

by Southern blot analysis. (See Chapter 5 for a description of Southern blot analysis.) DNA is cut with restriction endonucleases and probed with labeled *P* sequences to determine the number of insertions. Lines containing multiple insertions should be discarded because these lines will be difficult to analyze. G_3 lines with single inserts are then crossed to *Drosophila* stocks containing appropriate **balancer chromosomes**. Balancer chromosomes function to prevent crossing over between homologous chromosomes and thus help to maintain stable stocks. The location of the transposon in each single-insert line can be determined by *in situ* hybridization to salivary gland chromosomes.

Transformation success rates vary from experiment to experiment and experimenter to experimenter. Usually, it is important to obtain about 10 single-insert lines containing a transposon of interest. This may require microinjecting 600 or more embryos because survival of embryos after microinjection averages 30 to 70%, and of these, only 50 to 60% survive to adulthood (G_0). Even after G_0 adults are obtained, damage caused by microinjection may result in early death or sterility in 30 to 50%.

Transformation does not take place in all germ-line cells in an injected embryo. Usually only a small fraction of the germ-line cells of a G_0 individual produces transformed G_1 progeny. Thus, it is important to maximize the recovery of G_1 progeny from each G_0 individual injected to increase the probability of detecting progeny in which integration of *P* elements occurred. The size of the introduced *P* element is another factor that may influence transformation success; the larger the construct, the less frequent the insertion.

Detailed information on the life history and culture of *Drosophila* is available in a variety of references (Roberts 1986, Ashburner 1989, Matthews 1994, Horn and Wimmer 2000), as are detailed protocols for transforming *Drosophila* with *P* vectors (Karess 1987, Spradling 1986). The protocols provide complete information on the appropriate equipment for microinjection, and on how to stage and dechorionate embryos, align them on slides, desiccate them, and inject them in the region that contains the pole cells. Directions are available for preparing the DNA for injection and for pulling the very fine glass needles required.

P vectors have been engineered with different characteristics and functions (for examples, see Rubin and Spradling 1983, Karess and Rubin 1984, Cooley et al. 1988, Handler et al. 1993b, Horn and Wimmer 2000, Horn et al. 2000). Generally, the vectors contain restriction sites for cloning, and usually contain one or more selectable marker gene(s).

9.7.2. Characterizing Transformants

Identification of transformed flies is achieved in several ways. If a visible marker, such as an eye color, is included in the vector, then putatively transformed *D. melanogaster* can be determined visually.

Ideally, DNA from putatively transformed lines will be extracted and analyzed by Southern blot analysis to confirm the number of insertions in each line. If large numbers of fly lines need to be characterized, dot-blot analysis can be done. *In situ* hybridization of larval salivary gland chromosomes will allow a determination to be made of the number of insertions and their location(s). It is desirable to identify lines that carry only a single insertion if the timing and level of expression are to be determined. Different lines are likely to have different levels of expression because of position effects (Spradling and Rubin 1983, Levis et al. 1985).

9.8. Using *P*-Element Vectors

9.8.1. Transposon Tagging

The insertion of *P* into a gene allows the isolation and cloning of that gene *if* the altered gene results in altered phenotype in *D. melanogaster*. However, because many *P* strains contain 30 to 50 copies of *P*, transposon tagging should be carried out in *D. melanogaster* strains lacking endogenous *P* elements.

Transposon tagging relies on the development of two specially designed *P* vectors (Cooley et al. 1988). The goal is to introduce a *single P* into the germ line of flies *lacking P*. One vector, called "jumpstarter," encodes transposase and mobilizes a second vector, called "mutator," to transpose (Figure 9.4). The structure of the mutator element facilitates identifying and cloning genes because it carries two selectable markers.

9.8.2. Expressing Exogenous Genes

Genetic engineering techniques permit the expression of exogenous genes in a variety of organisms, and the availability of a transformation method for *Drosophila* makes it possible to express interesting genes in this insect. For example Rancourt et al. (1990) obtained expression in *D. melanogaster* of two antifreeze protein genes isolated from the Atlantic wolffish, *Anarhichas lupus*. The two genes were cloned into a *P* vector with *Drosophila* yolk protein gene promoters. These highly active promoters were expressed in *Drosophila* females shortly after eclosion and remained active for several weeks. Transformed adult *Drosophila* females produced 1.5 to 5 mg/ml of antifreeze protein in their hemolymph. The antifreeze activity of the purified protein was determined by measuring freezing-point depression and had full biological activity.

9.8.3. Evaluating Position Effects

"**Transposon jumping**" can be employed to move stably inserted *P* elements lacking transposase to other sites within the genome. This allows researchers to explore the effects of position on gene expression. To induce jumping, embryos from a transformed strain are injected with helper plasmids that transcribe transposase. The transposase interacts with the terminal repeats of the inserted *P*, causing it to transpose to a new site where the gene

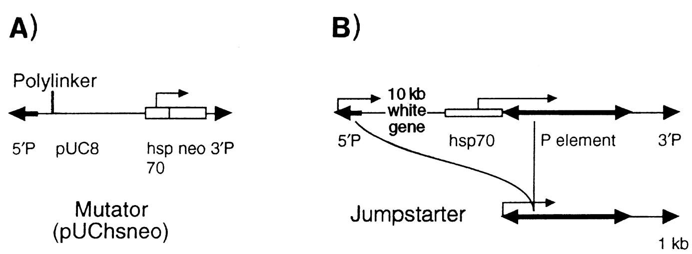


Figure 9.4. Two *P*-element vectors, mutator (A) and jumpstarter (B), were developed to facilitate insertion of a single *P* element to identify and clone genes in *D. melanogaster*. Jumpstarter encodes transposase and can therefore mobilize mutator. Mutator is able to transpose and carries ampicillin and neomycin resistance genes to facilitate identification and subsequent cloning of the *Drosophila* gene into which it has inserted (modified from Cooley et al. 1988).

located within the P vector experiences a new genomic environment. The helper element does not integrate, so the new strain will be stable until transposase is again supplied.

9.8.4. Targeted Gene Transfer

The ability to replace or modify genes in their normal chromosomal locations, **targeted gene transfer**, is a very valuable genetic tool (Ballinger and Benzer 1989, Kaiser and Goodwin 1990, Gloor et al. 1991, Sentry and Kaiser 1992, Lankenau 1995, Siegal and Hartl 1996, Golic et al. 1997, Rong and Golic 2000). Several methods have been evaluated to achieve such a goal.

The cut-and-paste mechanism of *P* transposition provided a model for inserting a gene into the gap left behind by a *P* (Figure 9.5). As noted above, *P* transposition leaves a ds gap in the original insertion site, and this gap may be repaired, using a template provided by a sister chromatid, by a homologous chromosome containing a homologous DNA sequence, or by an extrachromosomal element. If the sister chromatid or homologous chromosome has a second copy of the *P*, the *P* sequences will be restored in the gap, giving the impression that transposition has been replicative.

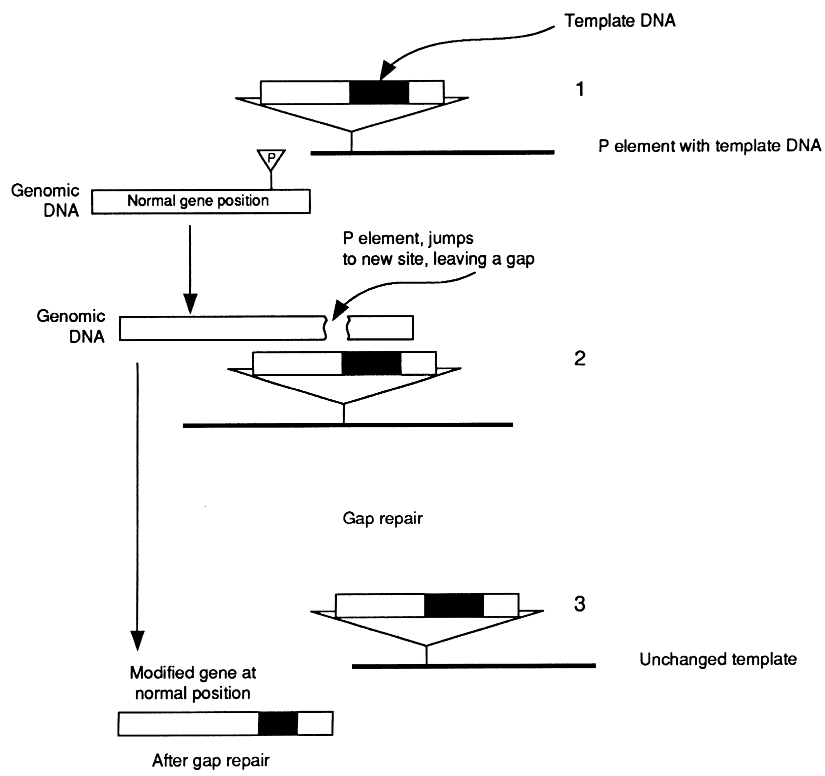


Figure 9.5. Targeted gene mutation in *D. melanogaster* is based on the gap repair hypothesis. If a *P* element jumps out of a normal gene, it will leave a gap that must be repaired. Repair is thought to involve using DNA with homologous ends from within the genome as a template for DNA repair. If a new *P* element with a modified gene structure is present, the sequence in the gap can be filled in using the modified gene as the template, leading to a targeted gene alteration (modified from Gloor et al. 1991).

Engels et al. (1990) proposed a method for **site-directed mutagenesis** (= targeted mutagenesis or targeted gene replacement). The first step is to insert a *P* into the gene of interest, preferably close to the site to be modified. This is feasible because many different colonies of *Drosophila* have been developed that contain *P* elements in known locations. The next step is to transfer the desired *replacement* gene into a second colony with a *P* vector. Then, individuals from the first colony are crossed with the second. A source of transposase is added to promote transposition and targeted gene replacement. In some cases, the replacement gene serves as the template to fill in the gap left when the *P* transposes. The result is that the repaired DNA is converted to the introduced sequence.

Targeted gene replacement (gene conversion) of the X-linked *white* locus was achieved by Gloor et al. (1991). They worked with a colony that carries a *P* in one of the *white* exons. This *P* cannot transpose because it lacks transposase. Transposase was added by crossing flies from this colony with a fly carrying an immobile *P* on chromosome 3 that can produce transposase. A vector, called P[walter], was injected that carried an altered *white* gene with 12-bp substitutions. These base-pair substitutions add or remove a restriction-enzyme site, which provides an efficient way to determine whether the replacement *white* was used to repair the gap induced by *P* transposition. About 1% of fly lines were identified with different amounts of gene conversions. Changes ranged from a few base pairs to alterations of at least 2790 bp.

A 1% gene conversion rate is sufficiently frequent to make targeted gene transfer a practical method for systematically altering genes in their normal locations to see how their function is modified. An advantage to targeted gene transfer is that it is possible to insert genes longer than 40 kb by this method. *P*-mediated transposition is limited to inserting DNA segments less than 40 kb in length.

Nassif and Engels (1993) investigated the length and stringency of homology required for repair of ds DNA breaks in *Drosophila* germ cells using the targeted gene transfer system. They found that a relatively short match (of a few hundred base pairs) of homologous sequence on either side of the target is sufficient to promote gap repair. However, the gap repair was sensitive to single base mismatches within the homologous regions. Interestingly, the data suggest that the ends of a broken chromosome can locate a single homologous template anywhere in the genome by using a short stretch of closely matching sequence. How this occurs remains mysterious, but the search is sufficiently efficient that up to several percent of the progeny exhibited gene conversion events (= targeted gene replacement) at the *white* locus. This high rate of gene conversion is considered to be unlikely if the process were dependent upon *random* collisions between homologous DNA sequences.

Transposable elements have been identified in most organisms investigated (Chapters 3, 4). If other TEs transpose by a mechanism similar to that of *P* elements, then targeted gene conversion could be feasible in other species of insects and mites (Sentry and Kaiser 1992). Targeted gene insertion would make it possible to introduce new genetic information into specific chromosomal sites, or to modify existing genes in directed ways. Random insertions can cause lethality if they insert into essential genes, or can result in poor levels of expression if insertion occurs into heterochromatic regions.

9.9. Transformation of Other Insects with *P* Vectors

DNA from *Drosophila melanogaster* has been introduced into *D. simulans* with *P*-element vectors. *P* transposition in *D. simulans* can produce a syndrome of hybrid dysgenesis similar

to that found in *D. melanogaster* (Daniels et al. 1989). *P* vectors also have transformed the more distantly related *D. hawaiiensis* (Brennan et al. 1984). Thus, *P* vectors can integrate and transpose in several *Drosophila* species.

Rio et al. (1988) suggested that *P* transposase was active in mammalian cells and yeast, which elicited optimism about the possibility of using *P* elements for genetically engineering other arthropods. Unfortunately, efforts to use *P* vectors to transform arthropod species outside the genus *Drosophila* have failed (O'Brochta and Handler 1988, Handler and O'Brochta 1991, Handler et al. 1993a, Handler and James 2000).

9.10. Evolution of Resistance to *P* Elements

The spread of *P* elements into populations of *D. melanogaster* has occurred worldwide during the past 50 years. This invasion has been remarkable because intact autonomous *P* elements can induce severe disadvantages in individuals in the newly invaded populations. If *P* elements invade a small population, that population usually is lost (Engels 1997). If evolution of repression systems (resistance to transposition) fails to occur quickly enough, the invaded populations go extinct (Corish et al. 1996).

In fact, several *P* repressor systems (resistance mechanisms) have been identified. The repressors are transmitted *either* cytoplasmically (maternally inherited) *or* through the nuclear genome, in which case the transmission is biparental. The repressor systems have been classified as P, M', or Q (Corish et al. 1996, Badge and Brookfield 1998, French et al. 1999).

P fly strains have a strong maternally inherited repression system called P cytotype (Engels 1989). **P cytotype** is mediated by a 66-kDa protein produced by differential splicing of the complete element's transcript (Laski et al. 1986). When P females are crossed to a strong P line, less than 10% of the ovaries are dysgenic, indicating that P strains strongly repress hybrid dysgenesis. If P males are crossed to M females (which lack a repression system), more than 90% of the ovaries are dysgenic in their progeny. P strains are strong inducers of transposition.

M' strains also contain repressor elements of *P*. Transposition repression in M' strains is due to the KP element (French et al. 1999). M' strain females display intermediate levels of repression of dysgenesis when crossed to P males. Both males and females from M' strains are able to pass the repressing factor to their progeny.

Q strains can strongly repress transposition and also display a low induction of transposition. Some Q strains show a maternal mode of inheritance of repression while others have biparental mode of inheritance. It is thought that a repressor (SR) results from a 309-bp deletion at the 3′ end of the *P* element. The SR repressor cannot produce functional transposase but can produce the 66-kDa repressor and a novel 75-kDa protein, both of which may be involved in Q type repression (French et al. 1999).

Evolution of resistance to *P* elements can develop rapidly, as demonstrated by two surveys of *D. melanogaster* along a 2900-km cline along the eastern coast of Australia. The first occurred in 1983 and the second in 1993. In 1983, P populations were found in the north, Q populations at central locations, and M' populations in the south (French et al. 1999). After 10 years, Q and M' populations had increased their range at the expense of P lines. Some of the northern Q lines transmitted repression through both sexes. French et al. (1999) speculated that the P and M' mechanisms of repression may be early, emergency responses to the harmful effects of transposition by *P*. The surviving *D. melanogaster* populations then

may have the opportunity to evolve a superior mechanism to improve fitness by acquiring the biparentally transmitted Q repression system.

In species of *Drosophila* in which *P* elements have been present for a long time, no complete functional *P* has been found (French et al. 1999). Instead, many populations contain tandem repeats of elements with degenerate fourth exons, which might encode some repressor activity. In *D. nebulosa*, a complete element was isolated, but the element contained many base changes in all four exons and was nonfunctional. These results reinforce the notion that active transposition of *P* is highly detrimental to species of *Drosophila* in the wild.

9.11. Using *P* to Drive Genes into Populations

There is interest in using TEs, such as *P*, as drivers for inserting engineered genes into natural populations for insect pest control. Some computer simulations and empirical studies have used *D. melanogaster* as a model system (Hastings 1994, Carareto et al. 1997). Several different computer simulations suggest that TEs may be used successfully to drive specific genes into pest populations, including populations with different sizes, reproductive rates, density dependence, and transposition frequency. Typically, an equilibrium was reached within 50 generations, especially if 5 or 10% of the population initially carried the TE. This topic will be discussed further in Chapter 14.

9.12. Relationship of *P* to Other Transposable Elements

The *Tc1* transposable element from the nematode *Caenorhabditis elegans* is part of a class of TEs that are structurally similar to the *P*, *HB*, and *Uhu* elements of *Drosophila* (Szekely et al. 1994). Both *P* and *Tc1* have perfect inverted repeats and contain ORFs encoding transposases. Both excise at a high frequency, and the mechanisms by which they do so appear similar (Szekely et al. 1994). Thus, it should be no surprise that *D. melanogaster* could be transformed with the *Tc1* element from *C. elegans* (Szekely et al. 1994), which supports the hypothesis that *Tc1*-like elements have a wide distribution within eukaryotes (Avancini et al. 1996).

The ability of the *Tc1* element to use *P* transposase suggests that TEs introduced into new species could be mobilized by endogenous native TEs. This raises the question as to whether genetically engineered insects transformed with a disabled TE could, in fact, be unstable because endogenous transposases could allow the inserted TE vector to move. This issue will have to be addressed in risk assessments conducted prior to the release of transgenic insects, even those transformed with disabled TE vectors, because "conversion" could occur (see below).

9.13. Other TEs Can Transform *D. melanogaster*

Several types of TEs have been used to transform *D. melanogaster*, including *piggyBac* (Lobo et al. 1999), *hobo* (Ladeveze et al. 1998), *mariner* (Garza et al. 1991), and *Minos* (Loukeris et al. 1995). *Hobo* occurs naturally in populations of *D. melanogaster*, so it is not surprising that it can serve as a transformation vector. *Minos* was discovered

in a related species, *D. hydei*. A TE vector derived from *mariner*, a TE found originally in *D. mauritiana*, also is effective in transforming *D. melanogaster* (Garza et al. 1991). *piggyBac*, isolated from a different order (Lepidoptera), is able to transpose in *D. melanogaster*. *piggyBac* originally was discovered located within a nuclear polyhedrosis virus infecting a *Trichoplusia ni* cell line. The finding of *piggyBac* within a virus suggests one mechanism by which TEs could move horizontally between different insects. *piggyBac* is related to Class II short inverted repeat elements, which includes *hobo*, *Minos*, *Hermes*, *mariner*, *P*, *Tc1* (found in nematodes), and *Ac* (found in plants).

9.14. Improved Transformation Tools for *Drosophila*

The efficiency of TE-mediated germ-line transformation is dependent on both the efficiency of the vector and the ability to detect (select) the transformed progeny.

Benedict et al. (1994) reported that a parathion hydrolase gene (*opd*), isolated from bacteria, under the control of a *Drosophila hsp70* promoter, produced good levels of resistance to paraoxon in *D. melanogaster*. Thus, this gene could serve as a semidominant selectable marker to detect transformation. Benedict et al. (1994, 1995) suggested the *opd* gene could be inserted into beneficial arthropods (parasitoids and predators) for improved pest management. The resistant natural enemies could survive the treatment with organophosphate insecticides, while the target pests could not. However, the use of organophosphate insecticides in the United States has declined because of their long residual activity and high toxicity to mammals, and these products are expected to be phased out of use as a result of the Food Quality Protection Act.

The use of the green fluorescent protein (GFP) as a selectable marker in transformed *D. melanogaster* has improved detection (Brand 1995, Yeh et al. 1995, Tsien 1998, Hazelrigg 2000). GFP also has been used as a selectable marker in other insects (Plautz et al. 1996, Pinkerton et al. 2000). Horn et al. (2000) reported that eye-specific expression of GFP outperforms the mini-*white* marker in *Drosophila* germ-line transformation experiments and recommended its use when transforming other arthropods.

An improved method of gene targeting involves the insertion of the FLP-FRT system into the chromosomes of *D. melanogaster* via *P*-mediated transformation (Golic et al. 1997). TE-mediated transformation is essentially a random event, and transgenes end up scattered throughout the genome in multiple copies, which results in position effect variation in expression and gene silencing (see below). The ability to target a transgene to a single site makes comparison of transgenes and their regulatory elements simpler because the various constructs can all be evaluated within a single chromosomal environment.

The FLP-FRT site-specific recombination system is based on a target site (FRT) and the FLP (site-specific recombinase enzyme) system of yeast (Kilby et al. 1993). This system has two 34-bp recombinase target sites arranged in an inverted orientation. The recombinase can excise the FRT-flanked DNA from the chromosome with nearly 100% efficiency, and the DNA is excised as an intact circular DNA that carries one FRT site. Thus, the FLP-mediated excision generates a single copy of the donor DNA in every cell as an extrachromosomal circle. If there is another FRT site (target site) elsewhere in the genome, when the extrachromosomal circle and chromosomal target come into contact FLP can mediate a second round of recombination that will integrate the circular DNA at the target site.

Another approach to gene targeting uses *D. melanogaster*'s endogenous DNA repair machinery and recombination to substitute one allele for another at a targeted gene or

to integrate DNA at a target site as determined by DNA sequence homology (Rong and Golic 2000). If this method proves of general use (Engels 2000), it will provide *Drosophila* geneticists with the ability to do "reverse genetics" for the first time. **Reverse genetics** is the induction of a mutation in a gene in order to determine its phenotype.

9.15. TE Vectors to Transform Insects Other Than *Drosophila*

The genetic modification of pest and beneficial insects by recombinant DNA methods to reduce their pest status or to improve their beneficial effects, respectively, is a goal of a growing number of scientists (Chapter 14). A mechanism with which to reliably insert exogenous DNA into the genome of the target insect and to have it transmitted in a stable manner in the germ line is required to achieve such goals. The *P* model has dominated the efforts of most scientists and TE insertion vectors have been developed from *piggyBac*, *mariner*, *Hermes*, *Minos*, and *hobo*.

9.15.1. *piggyBac*

Using a plasmid-based interplasmid transposition assay, the *piggyBac* vector was shown to transpose in a *Spodoptera frugiperda* cell line (Fraser et al. 1995, Elick et al. 1996), as well as embryos of *D. melanogaster*, the yellow fever mosquito *Aedes aegypti*, and the cabbage looper *Trichoplusia ni* (Lobo et al. 1999). The *piggyBac* vector has transformed two agricultural pests, the Mediterranean fruit fly *Ceratitis capitata* (Handler et al. 1998) and the pink bollworm *Pectinophora gossypiella* (Peloquin et al. 2000). It was used to transform the silkworm *Bombyx mori* (Tamura et al. 2000). *piggyBac* appears to have a broad host range.

Transformation rates with *piggyBac* average 2 to 5%. The insertions appear to be precise (producing a characteristic TTAA duplication at the insertion site) and stable, suggesting that this vector may be used to transform diverse insects (Fraser 2000).

9.15.2. *Hermes*

The *Hermes* TE was discovered in the housefly *Musca domestica* (O'Brochta et al. 1996). Cell lines of *Anopheles gambiae* were stably transformed by *Hermes* (Zhao and Eggleston 1998), and *Hermes* transposed in embryos of *Aedes aegypti* (Sarkar et al. 1997b). *Hermes* was shown to be functional in dipteran families (Drosophilidae, *D. melanogaster*; Calliphoridae, *Lucilia cuprina*; Tephritidae, *Ceratitis capitata* and *Bactrocera tryoni*; and Muscidae, *Musca domestica* and *Stomoxys calcitrans*) (Atkinson and O'Brochta 2000, O'Brochta et al. 2000).

Interestingly, the two strains of *M. domestica* tested exhibited the lowest rates of transformation (Sarkar et al. 1997a), perhaps because *Hermes* is endemic in this species and some form of resistance to *Hermes* has been selected for. Likewise, the *L. cuprina* strain tested exhibited low levels of transposition, perhaps because a *Hermes*-like element called *hermit* is present in the genome that elicited a partial resistance.

The mosquito *Aedes aegypti* (Jasinskiene et al. 1998, 2000) was transformed with *Hermes*. However, integrations of *Hermes* into *A. aegypti* did not occur precisely at the end of the terminal inverted repeats and were accompanied by small deletions in the plasmids.

These abnormal integrations also did not produce the typical 8-bp duplications at the insertion sites, suggesting that the vector could have integrated into the genome by general recombination or through a partial replicative transposition. As a result, Jasinskiene et al. (2000) concluded that this insertion mechanism by *Hermes* in *A. aegypti* "precludes its immediate use in experiments that involve field release of transformed animals into the field," although *Hermes* remains useful for laboratory experiments.

9.15.3. *Minos*

Minos can transpose in *Drosophila*, *Bombyx mori*, and *Anopheles stephensi* cells and embryos (Catteruccia et al. 2000a, Klinakis et al. 2000a, Shimizu et al. 2000) and also produce stable germ-line transformation (Catteruccia et al. 2000b).

Surprisingly, *Minos* can transform human cell lines, making it a useful tool for mutagenesis and functional analysis (Klinakis et al. 2000b). *Minos*'s ability to transform human cells suggests that it potentially has a very wide host range, which could elicit concerns about risks if *Minos* is used to transform insects destined for release into the field in pest management programs. To reduce potential risk with insects destined for field release, it may be necessary to eliminate the *Minos* element after transformation of the insect line.

9.15.4. *mariner*

This TE initially was isolated from *Drosophila mauritiana* but is extremely widespread among arthropods (Robertson 1995). The phylogeny of *mariner* elements isolated from diverse organisms is not congruent with their host species, indicating that *mariner* has been highly active in trans-order horizontal transfers (Robertson 1995, Robertson and Lampe 1995).

The host range of *mariner* is amazingly broad; *mariner* has been found in the insect-parasitic nematode *Heterorhabditis bacteriophora* (Grenier et al. 1999), the root-knot soil nematode *Meloidogyne* (Leroy et al. 2000), three flatworms (*Dugesia tigrina*, *Stylochus zebra*, *Bdelloura candida*) (Garcia-Fernandez et al. 1995), and two hydras (*Hydra littoralis* and *H. vulgaris*) (Robertson 1997). The host range of *mariner* extends to mammals (Auge-Gouillou et al. 1995, Oosumi et al. 1995, Robertson and Martos 1997). Auge-Gouillou et al. (1995) found that native *mariner* elements could be amplified by the PCR from human, mouse, rat, Chinese hamster, sheep and cow.

A *mariner* vector was used to transform the chicken (Sherman et al. 1998) and the zebrafish *Danio rerio* (Fadool et al. 1998). It even was used to transform the flagellate protozoan *Leishmania major*, which indicates that *mariner* has a general ability to "parasitize the eukaryotic genome" (Gueiros-Filho and Beverley 1997), perhaps because host proteins are not required for successful transposition (as they are if *P* is to transpose).

At least two different subfamilies of *mariner* have been isolated from the human genome, suggesting multiple horizontal transfers into the human genome have occurred, and Oosumi et al. (1995) suggested that *mariner* could be used as a transformation vector of humans. *A mariner* vector was genetically engineered to make it more active in humans; through genetic recombination and site-directed mutagenesis of a *mariner*-like defective element from fish, a new element called *Sleeping Beauty* was constructed that had 25-fold higher levels of activity in human cells than the "standard" *mariner* (Plasterk et al. 1999).

So far, all *mariner* elements discovered in humans are "molecular fossils derived from a *mariner* that was long ago active in the genome of a human ancestor," with each copy

having multiple mutations (Robertson and Martos 1997). Robertson and Zumpano (1997) found that *mariner* sequences are present in all major primate lineages and estimated that there are about 200 copies of one (*Hsmar1*) in the human genome, as well as approximately 2400 copies of a derived 80-bp inverted repeat structure and about 46,000 copies of single inverted repeats, suggesting that *mariner* had "a considerable mutagenic effect on past primate genomes."

The human genome is estimated to have been invaded by at least 14 families of TEs and is estimated to have more than 100,000 degenerate copies of TEs (Smit and Riggs 1996). These include elements called *pogo* (originally discovered in *Drosophila*) and *Tigger*, which are related to the *Tc1* and *mariner* TEs (Robertson 1996).

Despite the successes in transforming chickens, fish, and other organisms, rates of transformation of arthropods with *mariner* vectors have been low (Lampe et al. 2000). Coates et al. (1995) showed that *mariner* could excise in *D. melanogaster*, *D. mauritiana, Lucilia cuprina*, and *Bactrocera tryoni* embryos in excision assays. Wang et al. (2000) showed that *mariner* could mediate excision and transposition in *Bombyx mori* tissue culture cells. Later, *mariner* was shown to transpose in embryos of *Aedes aegypti* and to transform this vector of yellow fever with an eye-color gene (Coates et al. 1998). Mutants of the transposase gene from the *mariner* isolated from the horn fly *Haematobia irritans* have been screened, and some were found to have 4- to 50-fold increases in activity, indicating that *mariner* vectors could be developed that are more active in arthropods (Lampe et al. 1999).

9.15.5. *hobo*

The *hobo* vector transposed in a plasmid-based excision assay in several drosophilid species (Handler and Gomez 1995) and in cells of cabbage looper *Trichoplusia ni* and corn earworm *Helicoverpa zea* (DeVault et al. 1996), as well as in several tephritids, *Anastrepha suspensa*, *Bactrocera dorsalis*, *B. cucurbitae*, *C. capitata*, and *Toxotrypana curvicauda* (Handler and Gomez 1996). Elements related to *hobo* were found in most of these tephritids. Germ-line transformation of *Drosophila virilis* was mediated by *hobo* (Lozovskaya et al. 1996).

Excision of *hobo* from *H. zea* was stimulated by heat shocks that presumably stimulated the production of an endogenous *hobo*-like transposase. The excision rate was 8- to 10-fold higher than that seen for the normal host or other dipteran species (Atkinson et al. 1993) and, in hindsight, could have been predicted because *hobo* had been found previously in *H. zea* (DeVault and Narang 1994). The instability indicates the importance of checking the target insect species' genome to be sure that endogenous elements related to the TE vector are lacking before conducting transformation experiments.

9.16. Cross Mobilization of TE Vectors

Laboratory assays were conducted to compare the ability of *Minos*, *piggyBac*, *mariner*, and *Hermes* vectors to **cross mobilize** each other because these TEs have a wide array of family members (Sundararajan et al. 1999). The *hobo* transposase functioned equally well with *hobo* and *Hermes* substrates. On the other hand, the *Hermes* transposase rarely was able to excise the *hobo* elements from plasmids.

The hAT family of elements (which includes TEs from widely divergent taxa, including plants, fungi, fish, insects, and humans) appears able to function in novel hosts, and to move horizontally relatively easily (Kidwell and Lisch 1997, Kempken and Windhofer 2001).

These attributes make them especially desirable as vectors for inserting transgenes into arthropods, but could be considered negative attributes from the point of view of risk assessments when transgenic insects are being evaluated for release into the environment (Hoy 2000). The ability of different TEs to mobilize endemic (native) TEs (cross mobilization) is not limited to *Hermes* and *hobo* (Sundararajan et al. 1999).

9.17. Conversion of Inactive TE Vectors to Activity

The ability of disabled TE vectors to function in transgenic arthropods also should be evaluated before transgenic arthropods are released into the environment (Hoy 2000). An inactivated *P* vector was converted to activity in a process called **conversion**, through the interaction of three different *P* partners (Peronnet et al. 2000).

The defective *P* vector was converted into an active TE through a three-step process. The defective *P* (unable to move) used a remote template (another *P* that was itself unable to transpose because it lacked 21 bp at its 5′ end) for part of the template for the new element. The new element also had a restored 5′ end that allows it to transpose, which it obtained from a third element. This example provides strong evidence that the search for homology does occur during the DNA repair process after a ds break (Peronnet et al. 2000). Conversion to activity could, in some cases, make a transgene unstable within the transgenic insect's genome and could pose a potential risk for horizontal gene transfer.

9.18. Suppression of Transgene Expression

A variety of transgenic plants and mammals are known to inactivate multiple copies of genes if they produce an overexpression or abnormal transcription (Henikoff 1998). The inactivation phenomenon is thought to be due to the evolution of systems that prevent high levels of expression of TEs or of viruses. In fungi and plants, this **gene silencing** is associated with several mechanisms: methylation of the DNA, or posttranscriptional and transcriptional processes.

Transgene silencing has been described in *D. melanogaster* for the *white-alcohol dehydrogenase* transgenes (Pal-Bhadra et al. 1999). Transgene silencing in *Drosophila* also is associated with the production of heterochromatin (Dorer and Henikoff 1994, 1997). Methods to eliminate transgene silencing will be necessary, or this phenomenon could reduce the effectiveness of transgenic insects released for pest management programs (Hoy 2000).

9.19. Other Transformation Methods

In addition to the use of TE vectors, other transformation methods have been attempted for transforming insects, including the use of several types of viruses as vectors. Within the past 5 years dramatic advances have been made in developing reliable, stable methods for transforming arthropods other than *Drosophila* (Ashburner et al. 1998, Atkinson et al. 2001). Despite the ability to transform a variety of arthropod species, considerable work remains to be completed before transgenic arthropods can be used in practical pest management programs, as will be discussed in Chapter 14.

General References

Demerec, M., Ed. (1994). *Biology of* Drosophila. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. This is a facsimile edition of the original, first published by John Wiley and Sons in 1950.

Greenspan, R. J. (1997). *Fly Pushing. The Theory and Practice of* Drosophila *Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hartenstein, V. (1993). *Atlas of* Drosophila *Development*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sullivan, W., M. Ashburner, and R. S. Hawley (2000). Drosophila *Protocols*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

References Cited

Anxolabehere, D., and G. Periquet (1987). *P*-homologous sequences in Diptera are not restricted to the Drosophilidae family. *Genet. Iber.* **39**: 211–222.

Anxolabehere, D., M. G. Kidwell, and G. Periquet (1988). Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila *melanogaster* by mobile *P* elements. *Mol. Biol. Evol.* **5**: 252–269.

Ashburner, M. (1989). Drosophila. *A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ashburner, M., M. A. Hoy, and J. J. Peloquin (1998). Prospects for the genetic transformation of arthropods. *Insect Mol. Biol.* **7**: 201-213.

Atkinson, P. W., and D. A. O'Brochta (2000). *Hermes* and other *hAT* elements as gene vectors in insects, pp. 219–235. In: *Insect Transgenesis*, A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Atkinson, P. W., W. D. Warren, and D. A. O'Brochta (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90**: 9693–9697.

Atkinson, P. W., A. C. Pinkerton, and D. A. O'Brochta (2001). Genetic transformation systems in insects. *Annu. Rev. Entomol.* **46**: 317–346.

Auge-Gouillou, C., Y. Bigot, N. Pollet, M. H. Hamelin, M. Meunier-Rotival, and G. Periquet (1995). Human and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett.* **368**: 541–546.

Avancini, R. M. P., K. K. O. Walden, and H. M. Robertson (1996). The genomes of most animals have multiple members of the *Tc1* family of transposable elements. *Genetics* **98**: 131–140.

Badge, R. M., and J. F. Y. Brookfield (1998). A novel repressor of P element transposition in *Drosophila melanogaster*. *Genet. Res. Camb.* **71**: 21–30.

Ballinger, D. G., and S. Benzer (1989). Targeted gene mutations in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **86**: 9402–9406.

Benedict, M. Q., J. A. Scott, and A. F. Cockburn (1994). High-level expression of the bacterial *opd* gene in *Drosophila melanogaster*: improved inducible insecticide resistance. *Insect Mol. Biol.* **3**: 247–252.

Benedict, M. Q., C. E. Salazar, and F. H. Collins (1995). A new dominant selectable marker for genetic transformation; *Hsp70-opd*. *Insect Biochem. Mol. Biol.* **25**: 1061–1065.

Bingham, P. M., M. G. Kidwell, and G. M. Rubin (1982). The molecular basis of P-M hybrid dysgenesis: the role of the *P* element, a *P*-strain-specific transposon family. *Cell* **29**: 995–1004.

Brand, A. (1995). GFP in *Drosophila*. *Trends Genet.* **11**: 324–325.

Brennan, M. D., R. G. Rowan, and W. J. Dickinson (1984). Introduction of a functional *P* element into the germ-line of *Drosophila hawaiiensis*. *Cell* **38**: 147–151.

Carareto, C. M. A., W. Kim, M. F. Wojciechowski, P. O'Grady, A. V. Prokchorova, J. C. Silva, and M. G. Kidwell (1997). Testing transposable elements as genetic drive mechanisms using *Drosophila P* element constructs as a model system. *Genetica* **101**: 13–33.

Catteruccia, F., T. Nolan, C. Blass, H. M. Muller, A. Crisanti, F. C. Kafatos, and T. G. Loukeris (2000a). Toward *Anopheles* transformation: *Minos* element activity in anopheline cells and embryos. *Proc. Natl. Acad. Sci. USA* **97**: 2157–2162.

Catteruccia, F., T. Nolan, T. G. Loukeris, C. Blass, C. Savakis, F. C. Kafatos, and A. Crisanti (2000b). Stable germline transformation of the malaria mosquito *Anopheles stephensi*. *Nature* **405**: 959–962.

Clark, J. B., and M. G. Kidwell (1997). A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94**: 11428–11433.

Clark, J. B., P.C. Kim, and M. G. Kidwell (1998). Molecular evolution of *P* transposable elements in the genus *Drosophila*. III. The *melanogaster* species group. *Mol. Biol. Evol.* **15**: 746–755.

Coates, C. J., C. L. Turney, M. Frommer, D. A. O'Brochta, W. D. Warren, and P. W. Atkinson (1995). The transposable element *mariner* can excise in non-drosophilid insects. *Mol. Gen. Genet.* **249**: 246–252.

Coates, C. J., N. Jasinskiene, L. Miyashiro, and A. A. James (1998). *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti*. *Proc. Natl. Acad. Sci. USA* **95**: 3748–3751.

Cooley, L., R. Kelley, and A. Spradling (1988). Insertional mutagenesis of the *Drosophila* genome with single *P* elements. *Science* **239**: 1121–1128.

Corish, P., D. M. Black, D. W. Featherston, J. Merriam, and G. A. Dover (1996). Natural repressors of *P*-induced hybrid dysgenesis in *Drosophila melanogaster*: a model for repressor evolution. *Genet. Res. Camb.* **67**: 109–121.

Daniels, S. B., and L. D. Strausbaugh (1986). The distribution of *P* element sequences in *Drosophila*: the *willistoni* and *saltans* species groups. *J. Mol. Evol.* **23**: 138–148.

Daniels, S. B., L. D. Strausbaugh, L. Ehrman, and R. Armstrong (1984). Sequences homologous to *P* elements occur in *Drosophila paulistorum*. *Proc. Natl. Acad. Sci. USA* **81**: 6794–6797.

Daniels, S. B., A. Chovnick, and M. G. Kidwell (1989). Hybrid dysgenesis in *Drosophila simulans* lines transformed with autonomous *P* elements. *Genetics* **121**: 281–291.

DeVault, J. D., and S. K. Narang (1994). Transposable elements in Lepidoptera: *hobo*-like transposons in *Heliothis virescens* and *Helicoverpa zea*. *Biochem. Biophys. Res. Commun.* **203**: 169–175.

DeVault, J. D., K. J. Hughes, R. A. Leopold, O. A. Johnson, and S. K. Narang (1996). Gene transfer into corn earworm (*Helicoverpa zea*) embryos. *Genome Res.* **6**: 571–579.

Dorer, D. R., and S. Henikoff (1994). Expansions of transgene repeats cause heterochromatin formation ad gene silencing in *Drosophila*. *Cell* **77**: 993–1002.

Dorer, D. R., and S. Henikoff (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans*. *Genetics* **147**: 1181–1190.

Eggleston, W. B., D. M. Johnson-Schlitz, and W. R. Engels (1988). P-M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. *Nature* **331**: 368–370.

Elick, T. A., C. A. Bauser, and M. J. Fraser (1996). Excision of the *piggyBac* transposable element *in vitro* is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* **98**: 33–41.

Engels, W. R. (1989). P-elements in *D. melanogaster*, pp. 437–484. In: *Mobile DNA*. D. E. Berg, and M. M. Howe, Eds. Am. Soc. Microbiol., Washington, DC.

Engels, W. R. (1992). The origin of *P* elements in *Drosophila melanogaster*. *BioEssays* **14**: 681–686.

Engels, W. R. (1997). Invasions of *P* elements. *Genetics* **145**: 11–15.

Engels, W. R. (2000). Reversal of fortune for *Drosophila* geneticists. *Science* **288**: 1973–1974.

Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston, and J. Sved (1990). High-frequency *P* element loss in *Drosophila* is homolog dependent. *Cell* **62**: 515–525.

Fadool, J. M., D. L. Hartl, and J. E. Dowling (1998). Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. *Proc. Natl. Acad. Sci. USA* **95**: 5182–5186.

Fraser, M. J., Jr. (2000). The TTAA-specific family of transposable elements: identification, functional characterization and utility for transformation of insects, pp. 249–268. In: *Insect Transgenesis*, A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Fraser, M. J., L. Cary, K. Boonvisudhi, H. H. Wang (1995). Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **221**: 397–407.

French, D. J., P. Corish, M. Shi, and G. A. Dover (1999). Evolutionary flux of *P* element regulation in a *Drosophila melanogaster* hybrid dysgenesis cline. *Genet. Res. Camb.* **73**: 205–216.

Fujioka, M. J. B. Jaynes, A. Bejsovec, and M. Weir (2000). Production of transgenic *Drosophila*, pp. 353–363. In: *Methods in Molecular Biology*. Vol. (126). *Developmental Biology Protocols, Vol. II*. R. S. Tuan, and C. W. Lo, Eds. Humana Press, Totowa, NJ.

Garcia-Fernandez, J., J. R. Bayascas-Ramirez, G. Marfany, A. M. Munoz-Marmol, A. Casali, J. Baguna, and E. Slo (1995). High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthes): Evidence for a trans-phyla horizontal transfer. *Mol. Biol. Evol.* **12**: 421–431.

Garza, D., M. Medhora, A. Koga, and D. L. Hartl (1991). Introduction of the transposable element *mariner* into the germline of *Drosophila melanogaster*. *Genetics* **128**: 303–310.

Gloor, G. B., N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston and W. R. Engels (1991). Targeted gene replacement in *Drosophila* via *P* element-induced gap repair. *Science* **253**: 1110–1117.

Golic, M. M., Y. S. Rong, R. B. Petersen, S. L. Lindquist, and K. G. Golic (1997). FLP-mediated DNA mobilization to specific target sites in *Drosophila* chromosomes. *Nucleic Acids Res.* **25**: 3665–3671.

Grenier, E., M. Abadon, F. Brunet, P. Capy, and P. Abad (1999). A *mariner*-like transposable element in the insect parasite nematode *Heterorhabditis bacteriophora*. *J. Mol. Evol.* **48**: 328–336.

Gueiros-Filho, F. J., and S. M. Beverley (1997). Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. *Science* **276**: 1716–1719.

Hagemann, S., E. Haring, and W. Pinsker (1996). Repeated horizontal transfer of *P* transposons between *Scaptomyza pallida* and *Drosophila bifasciata*. *Genetica* **98**: 43–51.

Handler, A. M., and S. P. Gomez (1995). The *hobo* transposable element has transposase-dependent and -independent excision activity in drosophilid species. *Mol. Gen. Genet.* **247**: 399–408.

Handler, A. M., and S. P. Gomez (1996). The hobo transposable element excises, and has related elements in tephritid species. *Genetics* **143**: 1339–1347.

Handler, A. M., and A. A. James, Eds. (2000). *Insect Transgenesis. Methods, and Applications*. CRC Press, Boca Raton, FL.

Handler, A. M., and D. A. O'Brochta (1991). Prospects for gene transformation in insects. *Annu. Rev. Entomol.* **36**: 158–183.

Handler, A. M., S. P. Gomez, and D. A. O'Brochta (1993a). A functional analysis of the *P*-element gene-transfer vector in insects. *Arch. Insect Biochem. Physiol.* **22**: 373–384.

Handler, A. M., S. P. Gomez, and D. A. O'Brochta (1993b). Negative regulation of *P* element excision by the somatic product and terminal sequences of *P* in *Drosophila melanogaster*. *Mol. Gen. Genet.* **237**: 145–151.

Handler, A. M., S. D. McCombs, M. J. Fraser, and S. H. Saul (1998). The lepidopteran transposon vector, *piggyBac*, mediates germ-line transformation in the Mediterranean fruit fly. *Proc. Natl. Acad. Sci. USA* **95**: 7520–7523.

Hastings, I. M. (1994). Selfish DNA as a method of pest control. *Proc. Trans. R. Soc. Lond. B* **344**: 313–324.

Hazelrigg. T. (2000). GFP, and other reporters, pp. 313–343. In: Drosophila *Protocols*. W. Sullivan, M. Ashburner, and R. S. Hawley, Eds. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Henikoff, S. (1998). Conspiracy of silence among repeated transgenes. *BioEssays* **20**: 532–535.

Hickey, D. A. (1982). Selfish DNA: a sexually transmitted nuclear parasite. *Genetics* **101**: 519–535.

Horn, C., and E. A. Wimmer (2000). A versatile vector set for animal transgenesis. *Dev. Genes Evol.* **210**: 630–637.

Horn, C., B. Jannich, and E. A. Wimmer (2000). Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev. Genes Evol.* **210**: 623–629.

Houck, M. A. (1993). Mites as potential horizontal transfer vectors of eukaryotic mobile genes: *Proctolaelaps regalis* as a model, pp. 45–69. In: *Mites: Ecological and Evolutionary Analyses of Life-History Patterns*. M. A. Houck, Ed. Chapman and Hall, New York.

Houck, M. A., J. B. Clark, K. R. Peterson, and M. G. Kidwell (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **253**: 1125–1129.

Hoy, M. A. (2000). Deploying transgenic arthropods in pest management programs: Risks, and realities, pp. 335–337. In: A. M. Handler, and A. A. James, Eds., *Insect Transgenesis. Methods and Applications*. CRC Press, Boca Raton, FL.

Jasinskiene, N., C. J. Coates, M. Q. Benedict, A. J. Cornel, C. S. Rafferty, A. A. James, and F. H. Collins (1998). Stable transformation of the yellow fever mosquito, *Aedes aegypti*, with the *Hermes* element from the housefly. *Proc. Natl. Acad. Sci. USA* **95**: 3743–3747.

Jasinskiene, N., C. J. Coats, and A. A. James (2000). Structure of *Hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol.* **9**: 11–18.

Kaiser, K., and S. F. Goodwin (1990). "Site-selected" transposon mutagenesis of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **87**: 1686–1690.

Karess, R. E. (1987). *P* element mediated germline transformation of *Drosophila*, pp. 121–141. In: *DNA Cloning Volume II. A Practical Approach*. D. M. Glover, Ed. IRL Press, Oxford.

Karess, R. E., and G. M. Rubin (1984). Analysis of *P* transposable element functions in *Drosophila*. *Cell* **38**: 135–146.

Kempken, F., and F. Windhofer (2001). The hAT family: a versatile transposon group common to plants, fungi, animals,, and man. *Chromosoma* **110**: 1–9.

Kidwell, M. G. (1992). Horizontal transfer of *P* elements and other short inverted repeat transposons. *Genetica* **86**: 275–286.

Kidwell, M. G., and D. Lisch (1997). Transposable elements as sources of variations in animals and plants. *Proc. Natl. Acad. Sci. USA* **94**: 7704–7711.

Kilby, N. J., M. R. Snaith, and J. A. H. Murray (1993). Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9**: 413–421.

Klinakis, A. G., T. G. Loukeris, A. Pavlopoulos, and C. Savakis (2000a). Mobility assays confirm the broad host-range activity of the *Minos* transposable element and validate new transformation tools. *Insect Mol. Biol.* **9**: 269–275.

Klinakis, A. G., L. Zagoraiou, D. K. Vassilatis, and C. Savakis (2000b). Genome-wide insertional mutagenesis in human cells by the *Drosophila* mobile element *Minos*. *EMBO Rep.* **1**: 416–421.

Kobayashi, S., T. Kitamura, H. Sasaki, and M. Okada (1993). Two types of pole cells are present in the *Drosophila* embryo, one with and one without splicing activity for the third *P*-element intron. *Development* **117**: 885–893.

Ladeveze, V., I. Galindo, N. Chaminade, L. Pascual, G. Periquet and F. Lemeunier (1998). Transmission pattern of *hobo* transposable element in transgenic lines of *Drosophila melanogaster*. *Genet. Res. Camb.* **71**: 97–107.

Lampe, D. J., B. J. Akerley, E. J. Rubin, J. J. Mekalanos, and H. M. Robertson (1999). Hyperactive transposase mutants of the *Himar1 mariner* transposon. Proc. Natl. Acad. Sci. *USA* **96**: 11428–11433.

Lampe, D. J., K. K. O. Walden, J. M. Sherwood, and H. M. Robertson (2000). Genetic engineering of insects with *mariner* transposons, pp. 237–248. In: *Insect Transgenesis*. A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Lankenau, D. H. (1995). Genetics of genetics in *Drosophila: P* elements serving the study of homologous recombination, gene conversion, and targeting. *Chromosoma* **103**: 659–668.

Lansman, R. A., S. M. Stacey, T. A. Grigliatti, and H. W. Brock (1985). Sequences homologous to the *P* mobile element of *D. melanogaster* are widely distributed in the subgenus *Sophophora*. *Nature* **318**: 561–563.

Laski, F. A., D. C. Rio, and G. M. Rubin (1986). Tissue specificity of *Drosophila P* element transposition is regulated at the level of mRNA splicing. *Cell* **44**: 7–17.

Lee, S. H., J. B. Clark, and M. G. Kidwell (1999). A *P* element-homologous sequence in the house fly, *Musca domestica*. *Insect Mol. Biol.* **8**(**4**): 491–500.

Lemaitre, B., S. Ronsseray, and D. Coen (1993). Maternal repression of the *P* element promoter in the germline of *Drosophila melanogaster*: a model for the P cytotype. *Genetics* **135**: 149–160.

Leroy, H., F. Leroy, C. Auge-Gouillou, P. Castagnone-Sereno, F. Vanlerberghe-Masutti, Y. Bigot, and P. Abad (2000). Identification of *mariner*-like elements from the root-knot nematode *Meloidogyne* spp. *Mol. Biochem. Parasitol.* **107**: 181–190.

Levis, R., T. Hazelrigg, and G. M. Rubin (1985). Effects of genomic position on the expression of transduced copies of the *white* gene of *Drosophila*. *Science* **229**: 558–561.

Lobo, N., X. Li, and M. J. Fraser, Jr. (1999). Transposition of the *piggyBac* element in embryos of *Drosophila melanogaster, Aedes aegypti* and *Trichoplusia ni*. *Mol. Gen. Genet.* **261**: 803–810.

Loukeris, T. G., B. Arca, I. Livadaras, G. Dialektaki, and C. Savakis (1995). Introduction of the transposable element *Minos* into the germ line of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **92**: 9485–9489.

Lozovskaya, E. R., D. I. Nurminsky, D. L. Hartl, and D. T. Sullivan (1996). Germline transformation of *Drosophila virilis* mediated by the transposable element *hobo*. *Genetics* **142**: 173–177.

Matthews, K. A. (1994). Care, and feeding of *Drosophila*. *Methods Cell Biol.* **44**: 13–31.

Montchamp-Moreau, C., G. Periquet, and D. Anxolabehere (1991). Interspecific transfer of *P* elements by crosses between *Drosophila simulans* and *Drosophila mauritiana*. *J. Evol. Biol.* **4**: 131–140.

Nassif, N., and W. Engels (1993). DNA homology requirements for mitotic gap repair in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **90**: 1262–1266.

O'Brochta, D. A., and A. M. Handler (1988). Mobility of *P* elements in drosophilids, and nondrosophilids. *Proc. Natl. Acad. Sci. USA* **85**: 6052–6056.

O'Brochta, D. A., W. D. Warren, K. J. Saville, and P. W. Atkinson (1996). *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica*. *Genetics* **142**: 907–914.

O'Brochta, D. A., P. W. Atkinson, and M. J. Lehane (2000). Transformation of *Stomoxys calcitrans* with a *Hermes* gene vector. *Insect Mol. Biol.* **9**: 531–538.

Oosumi, T., W. R. Belknap, and B. Garlick (1995). *Mariner* transposons in humans. *Nature* **378**: 672.

Pal-Bhadra, M., U. Bhadra, and J. A. Birchler (1999). Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. *Cell* **99**: 35–46.

Peloquin, J. J., S. T. Thibault, R. Staten, and T. A. Miller (2000). Germ-line transformation of pink bollworm (Lepidoptera: Gelechiidae) mediated by *piggyBac* transposable element. *Insect Mol. Biol.* **9**: 323–333.

Peronnet, F., F. San Giorgioi, J. A. Lepesant, J. S. Deutsch and G. Gonzy-Treboul (2000). Three-partner conversion induced by the P-element transposase in *Drosophila melanogaster*. *Mol. Gen. Genet.* **262**: 1123–1131.

Petrov, D. A., J. L. Schutzman, D. L. Hartl, and E. R. Lozovskaya (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **93**: 8050–8054.

Pinkerton, A. C., K. Michael, D. A. O'Brochta, and P. W. Atkinson (2000). Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti*. *Insect Mol. Biol.* **9**: 1–10.

Plasterk, R. H. A., Z. Izsvak, and Z. Ivics (1999). Resident aliens. The Tc1/*mariner* superfamily of transposable elements. *Trends Genet.* **15**: 326–332.

Plautz, J. D., R. N. Day, G. M. Dailey, S. B. Welsh, J. C. Hall, S. Halpain, and S. A. Kay (1996). Green fluorescent protein and its derivatives as versatile markers for gene expression in living *Drosophila melanogaster*, plant, and mammalian cells. *Gene* **173**: 83–87.

Powell, J. R., and J. M. Gleason (1996). Codon usage and the origin of *P* elements. *Mol. Biol. Evol.* **13**: 278–279.

Rancourt, D. E., I. D. Peters, V. K. Walker, and P. L. Davies (1990). Wolffish antifreeze protein from transgenic *Drosophila*. *Bio/Technology* **8**: 453–457.

Rio, D. C., F. A. Laski, and G. M. Rubin (1986). Identification and immunochemical analysis of biologically active *Drosophila P* element transposase. *Cell* **44**: 21–32.

Rio, D. C., G. Barnes, F. A. Laski, J. Rine, and G. M. Rubin (1988). Evidence for *P* element transposase activity in mammalian cells, and yeast. *J. Mol. Biol.* **200**: 411–415.

Roberts, D. B., Ed. (1986). Drosophila. *A Practical Approach*. IRL Press, Oxford.

Robertson, H. M. (1995). The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.* **41**: 99–105.

Robertson, H. M. (1996). Members of the pogo superfamily of DNA-mediated transposons in the human genome. *Mol. Gen. Genet.* **252**: 761–766.

Robertson, H. M. (1997). Multiple *mariner* transposons in flatworms, and hydras are related to those of insects. *J. Hered.* **88**: 195–201.

Robertson, H. M., and D. J. Lampe (1995). Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **49**: 333–337.

Robertson, H. M., and R. Martos (1997). Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage. *Gene* **205**: 219–228.

Robertson, H. M., and K. L. Zumpano (1997). Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome. *Gene* **205**: 203–217.

Rong, Y. S., and K. G. Golic (2000). Gene targeting by homologous recombination in *Drosophila*. *Science* **288**: 2013–2018.

Rubin, G. M., and A. C. Spradling (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**: 348–353.

Rubin, G. M., and A. C. Spradling (1983). Vectors for *P* element-mediated gene transfer in *Drosophila*. *Nucleic Acids Res.* **11**: 6341–6351.

Rubin, G. M., M. G. Kidwell, and P. M. Bingham (1982). The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. *Cell* **29**: 987–994.

Sarkar, A., C. J. Coates, S. Whyard, U. Willhoeft, P. W. Atkinson and D. A. O'Brochta (1997a). The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica* **99**: 15–29.

Sarkar, A., K. Yardley, P. W. Atkinson, A. A. James, and D. A. O'Brochta (1997b). Transposition of the *Hermes* element in embryos of the vector mosquito, *Aedes aegypti*. Insect Biochem. *Mol. Biol.* **27**: 359–363.

Sentry, J. W., and K. Kaiser (1992). P element transposition and targeted manipulation of the *Drosophila* genome. *Trends Genet.* **8**: 329–311.

Sherman, A., A. Dawson, C. Mather, H. Gilhooley, Y. Li, R. Mitchell, D. Finnegan, and H. Sang (1998). Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nat. Biotechnol.* **16**: 1050–1053.

Shimizu, K., M. Kamba, H. Sonobe, T. Kanda, A. G. Klinakis, C. Savakis, and T. Tamura (2000). Extrachromosomal transposition of the transposable element *Minos* occurs in embryos of the silkworm *Bombyx mori*. *Insect Mol. Biol.* **9**: 277–281.

Siegal, M. L., and D. L. Hartl (1996). Transgene coplacement and high efficiency site-specific recombination with the Cre/*loxP* system in *Drosophila*. *Genetics* **144**: 715–726.

Simonelig, M., and D. Anxolabehere (1991). A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **88**: 6102–6106.

Smit, A. F. A., and A. D. Riggs (1996). *Tiggers* and other DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. USA.* **93**: 1443–1448.

Spradling, A. C. (1986). *P* element-mediated transformation, pp. 175–197. In: Drosophila. *A Practical Approach*. D. B. Roberts, Ed. IRL Press, Oxford, UK.

Spradling, A. C., and G. M. Rubin (1982). Transposition of cloned *P* elements into *Drosophila* germline chromosomes. *Science* **218**: 341–347.

Spradling, A. C., and G. M. Rubin (1983). The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* **34**: 47–57.

Spradling, A. C., D. Stern, I. Kiss, J. Roote, T. Laverty, and G. M. Rubin (1995). Gene disruptions using P elements: An integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. USA* **92**: 10824–10830.

Sundararajan, P., P. W. Atkinson, and D. A. O'Brochta (1999). Transposable element interactions in insects: crossmobilization of *hobo* and *Hermes*. *Insect Mol. Biol.* **8**: 359–368.

Szekely, A. A., R. C. Woodruff, and R. Mahendran (1994). *P* element mediated germ line transformation of *Drosophila melanogaster* with the *Tc1* transposable DNA element from *Caenorhabditis elegans*. *Genome* **37**: 356–366.

Taillebourg, E., and J. M. Dura (1999). A novel mechanism for *P* element homing in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **96**: 6856–6861.

Tamura, T., C. Thibert, C. Royer, T. Kanda, A. Eappen, M. Kamba, N. Komoto, J. L. Thomas, B. Mauchamp, G. Chavancy, P. Shirk, M. Fraser, J. C. Prudhomme, and P. Couble (2000). Germline transformation of the silkworm *Bombyx mori* L. using a *piggyBac* transposon-derived vector. *Nat. Biotechnol.* **18**: 81–84.

Tsien, R. Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* **67**: 509–544.

Wang, W., L. Swevers, and K. Iatrou (2000). *Mariner (Mos1)* transposase and genomic integration of foreign gene sequences in *Bombyx mori* cells. *Insect Mol. Biol.* **9**: 145–155.

Woodruff, R. C., J. N. Thompson, Jr., J. S. F. Barker, and H. Huai (1999). Transposable DNA elements, and life history traits: II. Transposition of *P* DNA elements in somatic cells reduces fitness, mating activity and locomotion of *Drosophila melanogaster*. *Genetica* **107**: 261–269.

Yeh, E., K. Gustafson, and G. L. Boulianne (1995). Green fluorescent protein as a vital marker and reporter of gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **92**: 7036–7040.

Zhao, Y. G., and P. Eggleston (1998). Stable transformation of an *Anopheles gambiae* cell line mediated by the *Hermes* mobile genetic element. *Insect Biochem. Mol. Biol.* **28**: 213–219.

10

Sex Determination in Insects

10.1. Overview

Resolving the molecular genetic basis of sex determination in arthropods has applications that potentially could result in improved genetic control programs for pest insects or in useful genetic modifications of beneficial species. Sex determination has been studied intensively in *D. melanogaster*. Sex determination in *D. melanogaster* has three components: 1) dosage compensation, 2) somatic sexual development, and 3) germ-line sexual development. The primary cue for determining sex in *D. melanogaster* is the number of X chromosomes relative to autosomes (A) in a cell (X:A ratio). This ratio determines somatic sex, germ-line sex, and dosage compensation by regulating functions of sets of regulatory genes. One model suggests that sex determination in all insects is based on modifications of the *Drosophila* scheme in which sexual development is controlled by a hierarchy of key regulatory genes. At the top of the regulatory cascade is *Sex lethal*$^{+}$ (*Sxl*$^{+}$), which must be ON to determine the female pathway. If *Sxl*$^{+}$ is OFF, the male pathway is the "default" developmental process. Sex subsequently is determined by the differential splicing of messenger RNAs. At the end of the pathway, sex determination is influenced by a DNA binding regulatory protein coded for by the *doublesex*$^{+}$ gene. Although sex determination in arthropods is clearly determined by chromosomal and genetic processes, environment also plays a role in some. Furthermore, infectious agents, including *Wolbachia*, *Rickettsia*, spiroplasmas, and viruses, are able to modify sex determination or sex ratio in many arthropods.

10.2. Introduction

Sexual reproduction results in genetic exchange, variation, and diversity. As a result of meiosis, genes obtained from different parents can be combined in a single descendant (review the discussion of meiosis in Chapter 4). New genotypes thus are constructed from preexisting variability by the mechanisms of **segregation** and **recombination** during meiosis. Homologous chromosomes separate (segregate) randomly to yield haploid gametes containing chromosomes derived from both the individual's maternal and paternal genomes. Recombination occurs during crossing over between paired homologous chromosomes during meiosis and results in new assortments of alleles.

Understanding sex determination and sex allocation in insects has both fundamental and applied applications. The evolutionary advantages of sexual reproduction and the reasons why organisms vary the sex ratio of their progeny are among the most discussed topics in evolutionary biology (for some samples, see Hamilton 1967, Hartl and Brown 1970, Maynard Smith 1978, Charnov 1982, Bull 1983, Thornhill and Alcock 1983, Lewis 1987, Michod and Levin 1988, Hamilton et al. 1990, Wrensch and Ebbert 1993, Crow 1994, Barton and Charlesworth 1998, Marin and Baker 1998, Partridge and Hurst 1998,

Werren and Beukeboom 1998, Antolin 1999, Keightley and Eyre-Walker 2000, West et al. 2000, Rice and Chippindale 2001).

10.3. Costs and Benefits of Sexual Reproduction

Theories about the evolution of sex have focused on the advantages of the combined effects of segregation and recombination. In a sexual population, advantageous mutations that arise at two different loci in two parents can be combined in one individual in later generations.

10.3.1. Sexual Reproduction Has Costs

Despite the advantages of sex, sex has costs, and Crow (1994) summarized these as follows: 1) sex expends energy that could be used for other purposes; 2) males are expensive, with 50% savings possible if males were eliminated; 3) sexual selection in sexual species often leads to maladapted traits and destructive competition for mates; 4) sexual species do not allow the perpetuation or fixation of novel genetic types because heterozygotes often are broken up by segregation and recombination, and changes in ploidy (such as triploids and aneuploids) cannot go through meiosis successfully; 5) sexual species have to find a mate, which can be a disadvantage in sparse populations, or species with limited motility, or colonizers of a new area; 6) sexual species are prone to sexually transmitted diseases and harmful transposons whose spread is facilitated by biparental inheritance; 7) short-term selection is slower in sexual than asexual species; 8) sexual species cannot colonize microhabitats without the distinctive properties of these adapted colonies being swamped by hybridization.

10.3.2. Advantages of Sex Must Be Large

A number of theories attempt to explain why sexual reproduction persists (Crow 1994). 1) One possibility is that sex provides an ability to **incorporate and accumulate favorable mutations**. Mutations that arise in an asexual species in different individuals cannot be combined in one individual easily; successive advantageous mutations would have to occur in the same asexual lineage, one after the other. 2) Sexual reproduction may allow the accumulation of favorable mutations when deleterious mutations are present (whereas the maintenance of a favorable mutation in an asexual population is dependent upon the relative fitness of the individual in which the mutation occurs). Thus, the value of sex "lies more in the ability to reassort existing genes as the environment changes and in the elimination of harmful mutations" (Crow 1994).

Sexual reproduction allows harmful mutations to be eliminated. This effect is based on a concept termed **Muller's ratchet**. Muller (1964) noted that in an asexual population, unless it is very large, it is unlikely that any individual is free of harmful mutations. In such a population, the most fit individual is one that has only one mutation. In the next generation, mutations occur again, and this time the most fit individual may have one new mutation, or two in total, and the "ratchet" has turned another cog. In the absence of reverse mutation, such a population would accumulate more and more deleterious mutations. In a sexual population, a mutant-free type can be created by recombination.

Sexual reproduction reduces the mutation load. The deleterious effects of mutations are related to their frequency of occurrence, not to the magnitude of their effects. The smaller

the effect of a mutation, the more individuals it will affect before it is eliminated from the population. Over time, the number of mutations removed per eliminated individual is much larger in a sexual than in an asexual population and the mutation load is reduced correspondingly.

Sexual reproduction and diploidy have evolutionary advantages by providing protection from somatic mutations (Crow 1994). Diploidy is common in higher organisms, which have the most extensive and highly differentiated soma. Each generation somatic tissues develop and are identical, except for somatic mutations which are often recessive. Diploidy thus diminishes the deleterious effects of recessive somatic mutations, which could destroy essential cells or initiate abnormal growth. Diploidy requires passing through a single cell each generation so the soma can begin anew without these mutations. Diploidy also permits the efficient repair of double-stranded breaks in DNA, which repairs mutations.

10.3.3. Origin of Sex

The reasons for maintaining sexual reproduction in current populations are "likely to be quite different from the mechanisms by which sex got started in the first place" (Crow 1994). It is generally accepted that sex was determined initially by an allelic difference at a gene located on a homologous pair of autosomes (Rice 1994, Lucchesi 1999). The two sexes thus consisted of individuals heterozygous or homozygous at this sex-determining locus. The transformation of autosomes bearing the sex-determining gene into heteromorphic sex chromosomes (such as X and Y) is thought to have occurred by the accumulation of mutations in the neighborhood of the sex-limited allele. The retention of such mutations is thought to be facilitated by a reduction in the rate of recombination in the chromosome of the individuals bearing the sex-limited allele.

Understanding the mechanisms of sex determination in insects provides insights into the regulation of development of a significant character in eukaryotes. Such knowledge could provide useful tools for the genetic improvement of arthropod natural enemies of pest arthropods and weeds and for genetic modification of pests, which could improve the methods by which genetic control programs are achieved (Shirk et al. 1988, Stouthamer et al. 1992, LaChance 1979, Grenier et al. 1998, Heinrich and Scott 2000, Robinson and Franz 2000). See also Chapter 14 for a discussion of genetic manipulation of pest and beneficial arthropods.

10.4. Sex Determination Involves Soma and Germ-Line Tissues

Sex determination involves both the soma and germ-line tissues (ovaries and testes). Sexual dimorphism in adult insects is often extreme, with differences in setal patterns, pigmentation, external genitalia, internal reproductive systems, and behavioral patterns (Greenspan and Ferveur 2000, Kopp et al. 2000).

How do sexually determined differences in the soma and germ line arise? The details are becoming clear for *D. melanogaster*, and some information is available for a few other economically important insects such as mosquitoes.

First, we will review the basic sex determination system in *D. melanogaster*. Then, sex determination in some other insects will be described. Finally, examples will be provided that illustrate the importance of extrachromosomal and microbial genes in modifying sex in many arthropods.

10.5. Sex Determination in *Drosophila melanogaster*

Developing an understanding of sex determination in *D. melanogaster* has relied on identifying a relatively few spontaneous mutants (Table 10.1), which suggests that the number of genes involved is relatively low (Belote et al. 1985, Slee and Bownes 1990, Cline and Meyer 1996). Sex determination in *D. melanogaster* involves three major components: dosage compensation, somatic cell differentiation, and germ-line differentiation. More is known about dosage compensation and somatic cell differentiation than about germ-line differentiation.

10.5.1. Dosage Compensation of X Chromosomes

A basic aspect of sex determination in insects with an XY sex determining system is dosage compensation of the X chromosomes (Baker et al. 1994). The mechanism of dosage compensation varies in different arthropods. **Dosage compensation** equalizes the amount

Table 10.1. Some Genes Involved in Somatic Sex Determination and Dosage Compensation in *D. melanogaster*

Gene	Function
Maternal genes	
daughterless$^+$	Necessary for numerator genes to act appropriately;
hermaphrodite$^+$	*da*$^+$ and *her*$^+$ activate *Sxl*$^+$ in female embryos.
extramachrochaetae$^+$	*emc*$^+$ and *gro*$^+$ negatively regulate *Sxl*$^+$ in female embryos.
groucho$^+$	
Numerator genes	
sisterless-A$^+$	Communicate X-chromosome dose in dosage compensation.
sisterless-B$^+$	X-linked genes involved in activating *Sxl*$^+$ in females;
sisterless-C$^+$	they "count" the number of X chromosomes and turn on
runt$^+$	*Sex-lethal*$^+$.
Zygotic genes	
Sex-lethal$^+$	Major control gene; produces a full-length protein in females; no protein produced in males.
transformer$^+$	Active, with *tra-2*$^+$; in regulating *dsx*$^+$ in females.
transformer-2$^+$	Active in females to induce female-specific *dsx*$^+$ expression and repress male-specific *dsx*$^+$ expression. Needed for spermatogenesis.
doublesex$^+$	Active in males to repress female differentiation; in females *dsx*$^+$ represses male differentiation; loss of function mutants result in intersexes in both males and females; a pivotal terminal differentiation switch.
intersex$^+$	Active in females with *dsx*$^+$ product to repress male differentiation; not needed in males.
fruitless$^+$	Gene product is necessary in nervous system of males to elicit normal mating behavior and development of male muscle (muscle of Lawrence).
male-specific lethal$^+$	All four genes regulate X chromosome transcription in males;
msl-1$^+$	*msl-2*$^+$ has no function in females. Absence of the MSL-2
msl-2$^+$	protein in females prevents formation of the compensasome.
msl-3$^+$	

Derived from Bownes (1992), Cline and Meyer (1996), Li and Baker (1998), Marin and Baker (1998).

of gene products produced by individuals with an XX/XY genetic system (males containing one and females two X chromosomes). Dosage compensation in *D. melanogaster* is achieved by **hypertranscription** of the single X chromosome in males (Marin et al. 2000). As a result, males produce *equivalent* amounts of gene product compared to females that have two X chromosomes. By contrast, dosage compensation in the mole cricket *Gryllotalpa fossor* is analogous to that in mammals; one of the two X chromosomes in females is transcriptionally **inactivated**, with the inactivation occurring randomly within each cell (Rao and Padmaja 1992). *G. fossor* males are XO and females are XX, and one of the two X chromosomes in female cells is late-replicating and transcriptionally silent.

Males (XY:AA) are **aneuploid** for an X, which is a large fraction of the total genome (the Y has only a few genes on it). Aneuploidy (when the chromosomal composition in a cell is not an exact multiple of the haploid set) is normally lethal to an organism.

In *Drosophila* males, hypertranscription of the single X chromosome requires the functions of autosomal genes, *male-specific lethal genes*$^{+}$ (*msl*$^{+}$), which are under the control of *Sxl*$^{+}$ (Table 10.1, Figure 10.1) and some RNAs that are associated with the chromatin (Kelley and Kuroda 2000). The MSL proteins are assembled with the RNAs in a remodeling complex (called a compensasome) on about 100 sites on the X chromosome in males (Figure 10.1). The resulting histone H4 acetylation leads to hypertranscription (Marin et al. 2000,

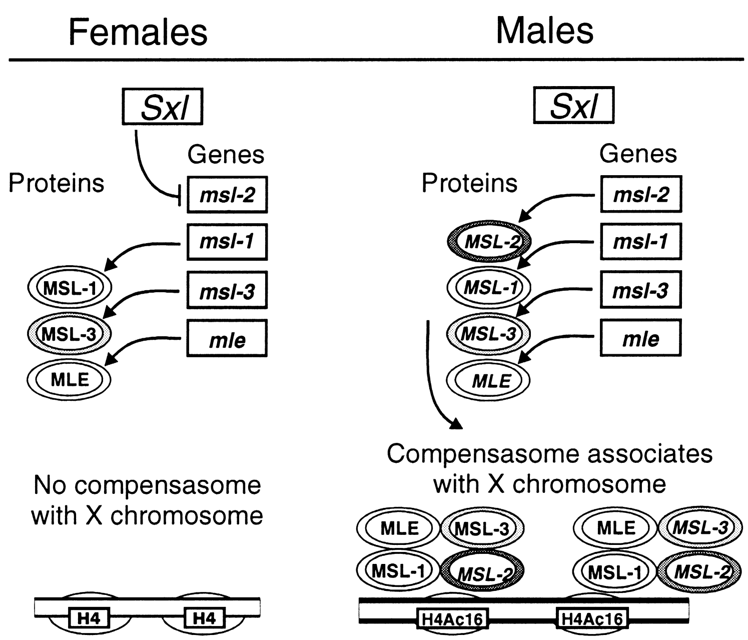


Figure 10.1. Model for the regulation of dosage compensation of X chromosomes in *D. melanogaster* males. MLE, MSL-l, and MSL-3 proteins are produced in both sexes. The *Sxl*$^{+}$ gene negatively regulates another gene so that no functional protein, probably MSL-2, is made in females. In the absence of this protein, the MLE, MSL-1, and MSL-3 proteins cannot associate stably as a "compensasome" with the X chromosome in females. The compensasome does associate with sites on the X chromosome in males that have the histone H4 acetylated at the lysine 16. As a result, the sole X chromosome is hypertranscribed in males.

Smith et al. 2000). By contrast, SXL protein in females prevents removal of a female-specific intron in the *msl-2*$^+$ mRNA; without this MSL-2 protein the other MSL proteins fail to assemble on the X chromosome and hypertranscription is prevented.

10.5.2. Somatic Sex Determination

The relative number of X chromosomes and autosomes in *D. melanogaster* is responsible for the primary step in sex determination immediately after fertilization. Cells with two X chromosomes and two sets of autosomes (2X:2A, or a ratio of 1.0) develop into females, while diploid cells (with 1X:2A, or a ratio of 0.5) develop into males (Figure 10.2). Flies with equal numbers of X chromosomes and autosomes (XX:AA, XXX:AAA, or X:A, or a ratio of 1.0) develop as females. Flies with an intermediate X:A ratio (XX:AAA) develop as intersexual flies that appear to be mosaics of discrete patches of male or female tissues. Distinct boundaries between cells exist in *Drosophila*, and sex determination takes place in individual cells (**cell autonomous sex determination**). Cell autonomous determination of

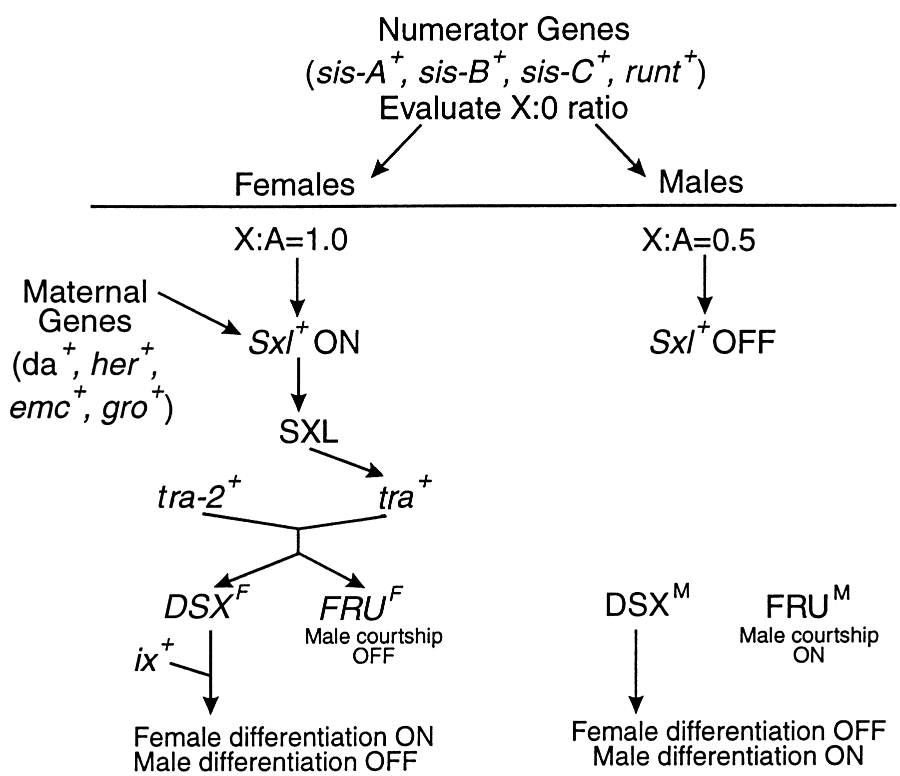


Figure 10.2. The general features of somatic sex determination in *D. melanogaster*. The ratio of X chromosomes to autosomes (A) determines whether *Sxl*$^+$ is ON. *Sxl*$^+$ produces a protein, SXL, that acts as a splicing factor on the RNA produced by the *tra*$^+$ gene, resulting in the production of active TRA protein. TRA, together with the product of the *tra-2*$^+$ gene, determine the female-specific splicing of the *dsx*$^+$ and *fruitless*$^+$ RNAs, which results in a cascade of genes functioning to produce a female. If *Sxl*$^+$ is OFF, the individual becomes a male because male-specific products (DSX^M and FRU^M) of the *dsx*$^+$ and *fru*$^+$ genes are produced.

sex has been assumed to be due to the lack of **sex hormones** in insects. However, DeLoof and Huybrechts (1998) note that hormones are involved in the development of secondary sexual characters in some insects.

The Y chromosome does not determine sex in *D. melanogaster*, although it is required for normal spermatogenesis and fertility. At least six genes on the Y chromosome are important in fertility, each performing a single, unique function (Hennig 1993, Hochstenbach et al. 1994). The fertility genes are greater than 1000 kb each and are highly susceptible to mutations (Hackstein and Hochstenbach 1995). The large amount of noncoding repetitive DNA on the Y appears to be selfish DNA because it has been impossible to assign any essential function to it (Hackstein and Hochstenbach 1995). The Y chromosome is important in sex determination in some insects (Marin and Baker 1998).

Numerator genes communicate the relative number or ratio of X chromosomes and autosomes. Numerator genes in *D. melanogaster* include *sisterless-A*$^+$ (*sis-A*$^+$), *sisterless-B*$^+$ (*sis-B*$^+$), *sisterless-C*$^+$ (*Sis-C*$^+$), and *runt*$^+$ (Cline and Meyer 1996). Once the X:A ratio is assessed, the activities of a relatively small number of major regulatory genes are triggered that ultimately lead to male or female differentiation in the soma (Figure 10.2). Somatic sexual differentiation is regulated through a cascade of sex-specific events in which RNA transcripts are differentially processed in males and females (Figure 10.2, Table 10.1). Note that in Figure 10.2 gene products are capitalized (SXL) while the gene (*Sxl*$^+$) is italicized.

Sex-lethal$^+$ is a key switch gene that, very early in development, affects both somatic sexual differentiation and dosage compensation (Figure 10.2). *Sex lethal*$^+$ codes for an RNA splicing enzyme. Its action on the next gene in the cascade, *transformer*$^+$, is restricted to females in its role in sexual differentiation. *Sex-lethal*$^+$ must be ON in females and OFF in males (Figure 10.2). Once the X:A ratio is read and the *Sxl*$^+$ gene is turned ON or OFF early in embryonic development, the developmental path chosen is stable (Cline and Meyer 1996).

Sxl$^+$ is transcribed in females in a complex manner. Two different promoters function in somatic cells; one (the establishment promoter, Sxl^{Pe}) acts very early and only for a brief period during nuclear cycle 12 to early cycle 14, ending when somatic cells first form in the young embryo. As this promoter shuts off, the second promoter (Sxl^{Pm}) comes on in both sexes. However, because the transcripts from this promoter require full-length SXL protein to remove a male-specific exon, only the expression of the Sxl^{Pm} in females generates mRNAs that encode full-length SXL protein. Thus, the earliest *Sxl*$^+$ transcripts differ from later transcripts, and male transcripts are inactive because they include an extra exon that stops the translation process. Initiation of *Sxl*$^+$ expression requires the action of genes from the mother (maternal genes such as *da*$^+$, *her*$^+$, *emc*$^+$, *gro*$^+$) (Figure 10.2). In males, the *Sxl*$^+$ master gene is OFF, and the four *male-specific lethal*$^+$ autosomal genes are ON, a combination which leads to male somatic sexual differentiation and hypertranscription of the single X chromosome (Figure 10.1).

Maternal genes also influence the development of progeny. Maternal genes function in one of two ways: either the mother produces a gene product that is transferred to and stored in the egg, or the mother's messenger RNA is transferred to and stored in the egg and subsequently is translated by the embryo. At least four maternal X:A signal transduction genes have been found, including *daughterless*$^+$ (*da*$^+$), *hermaphrodite*$^+$ (*her*$^+$), *extramachrochaetae*$^+$ (*emc*$^+$), and *groucho*$^+$ (*gro*$^+$) (Figure 10.1, Table 10.2). Female progeny of mothers with mutant forms of *da*$^+$ fail to activate the key master gene *Sxl*$^+$ and die as embryos. Male progeny of *da* mothers survive because they do not require *Sxl*$^+$.

Table 10.2. Examples of Nonnuclear Influences on Sex Determination or Sex Ratio in Arthropods[a]

Microorganism	Arthropod species (order)	Effect References
Arsenophonus nasoniae, gram-negative bacterium	*Nasonia vitripennis* (Hymenoptera)	Kills male eggs Gherna et al. 1991
Bacterial male-killing in Coleoptera		
Several, including: *Rickettsia* *Spiroplasma* *Wolbachia* (two types)	*Adalia bipunctata*	Males killed Werren et al. 1994 Hurst et al. 1999a Hurst et al. 1999b
Flavobacterium	*Coleomegilla maculata*	Hurst et al. 1997
Rickettsia	*Adalia decempunctata*	von der Schulenburg et al. 2001
Spiroplasma	*Harmonia axyridis*	Majerus et al. 1999
Wolbachia	*Tribolium madens*	Stevens 1993, Fialho and Stevens 2000
Rickettsia	*Brachys tessellates*	Lawson et al. 2001
Maternal sex ratio (MSR)	*Nasonia vitripennis* (Hymenoptera)	Results in nearly all female progeny Beukeboom and Werren 1992
Paternally transmitted sex ratio factor (PSR), a supernumerary B chromosome that is mostly heterochromatic	*Nasonia vitripennis* (Hymenoptera)	Fertilized eggs lose paternal chromosomes; females converted to males; transmitted only via sperm Werren et al. 1987, Nur et al. 1988, Beukeboom and Werren 1993
	Encarsia pergandiella? (Hymenoptera)	Mechanism speculative Hunter et al. 1993
Sex-ratio condition viral?	*Drosophila bifasciata* (Diptera)	Death of male embryos Leventhal 1968
Sex-ratio condition spiroplasmas	*D. willistoni* and related neotropical species (Diptera)	Males die as embryos Ebbert 1991
Sex-ratio condition	*Oncopeltus fasciatus* (Hemiptera)	Leslie 1984
Sterility in male progeny streptococcal L-form bacteria	*Drosophila paulistorum* (Diptera)	Induction of semispecies? Somerson et al. 1984
Thelytoky *Wolbachia*	*Encarsia formosa* (Hymenoptera)	Males produced after antibiotic treatment; microorganisms restore diploidy to unfertilized eggs, resulting in all female progeny (thelytoky) Zchori-Fein et al. 1992
Bacterium (EB)	*Encarsia pergandiella*	Bacterium unrelated to *Wolbachia* causes thelytoky and alters host selection behavior Zchori-Fein et al. 2001
Thelytoky lost maternally inherited *Wolbachia*	*Trichogramma* species (Hymenoptera)	Cure with antibiotics results in bisexual (arrhenotokous) populations Stouthamer et al. 1990, Stouthamer and Werren 1993, Stouthamer and Luck 1993

continues

Table 10.2. continued

Microorganism	Arthropod species (order)	Effect References
Wolbachia	*Metaseiulus occidentalis* (Acari)	Females eliminated in incompatible crosses Johanowicz and Hoy 1996, 1998
	Leptopilina hetrotoma (Hymenoptera)	Triinfected strain males crossed with uninfected or monoinfected females result in killed female eggs; monoinfected males crossed with uninfected females result in reduced progeny and more males Vavre et al. 2000

[a] *Wolbachia* also causes cytoplasmic incompatibility.

Sxl$^+$ is the master switch gene involved in both sex determination and dosage compensation. It regulates pre-mRNA splicing for itself and for *transformer*$^+$ (*tra*$^+$) and *male-specific-lethal-2*$^+$ (*msl-2*$^+$). Once *Sxl*$^+$ is ON in females, a second series of regulatory genes are important in differentiating between the alternative pathways in somatic cell development. These secondary switch genes include *transformer*$^+$ (*tra*$^+$), *transformer-2*$^+$ (*tra-2*$^+$), *intersex*$^+$ (*ix*$^+$), *doublesex*$^+$ (*dsx*$^+$), and *fruitless* (*fru*$^+$) (Figure 10.2). Mutations of *tra*$^+$, *tra-2*$^+$, and *ix*$^+$ affect somatic sex determination in females, but are not needed for male somatic differentiation (Table 10.1). In the absence of TRA proteins in males, the *fruitless*$^+$ gene transcript affects as many as 500 neurons in the brain (Figure 10.2), which regulates male sexual behavior and also affects the male-specific muscle (muscle of Lawrence, MOL) used in mating. Although *tra-2*$^+$ is not needed for male differentiation, it is critical for normal spermatogenesis in males.

The *doublesex*$^+$ locus is needed for differentiation of both male and female external morphology (Figure 10.2). The *dsx*$^+$ gene is a **double switch**, with only one switch functioning in a particular sex. When *dsx*$^+$ is active in males (producing the male gene product, DSX^M), it represses female differentiation. When *doublesex*$^+$ is active in females (producing DSX^F), and the *intersex*$^+$ gene product is present, male development is suppressed. If *dsx*$^+$ is inactivated, both male and female genes are active within a cell, which results in an intersexual phenotype at the cellular level.

The determination of sex during embryogenesis in *D. melanogaster* is transmitted through a hierarchy of regulatory genes to the terminal differentiation genes, whose products are responsible for the sexually dimorphic traits of the adult fly (Bownes 1992). The different activities of the regulatory genes in males and females are largely due to sex-specific differences in RNA splicing that lead to the production of functionally different transcripts in the two sexes (Baker 1989). The individual genes in this regulatory hierarchy not only are themselves controlled at the level of RNA splicing but, in turn, specify the splicing pattern of the transcripts of genes downstream in the hierarchy, producing a cascade of RNA splicing reactions. Thus, RNA processing is an important regulatory mechanism in this significant developmental pathway.

In addition to the sex-determination genes, there are genes whose products are responsible for the structure and function of sexually dimorphic somatic tissues (Kopp et al. 2000). A number of structural genes are controlled by the sex-determination regulatory pathway,

such as the yolk polypeptide genes which are expressed in the fat body in a female-specific manner.

10.5.3. Germ-Line Determination

Sex determination in the development of germ-line tissues in *D. melanogaster* is different from that in the soma (Pauli and Mahowald 1990, Janzer and Steinmann-Zwicky 2001, Vincent et al. 2001). Pole cells in the embryo are segregated into the posterior pole of the insect embryo before blastoderm formation and include the progenitors (stem cells) of the germ cells (Xie and Spradling 2000).

Components of the germ plasm (= pole plasm) are synthesized in the mother during oogenesis by a cluster of 15 nurse cells, which are connected to the oocyte at its anterior by cytoplasmic bridges. Pole plasm components are transported into the oocyte and translocated to the posterior pole of the egg. Maternally active genes important in the production of pole cells include *cappuccino*$^+$, *spire*$^+$, *staufen*$^+$, *oskar*$^+$, *vasa*$^+$, *valois*$^+$, *mago nashi*$^+$, and *tudor*$^+$ (Ephrussi and Lehmann 1992). These genes also are important in the formation of normal abdomens in *Drosophila*.

During embryogenesis, prospective male and female germ cells are indistinguishable, but differentiation is begun during the larval stage, when male gonads grow larger than female gonads because they contain more germ cells. The sexual identity of germ cells is determined by both the X:A ratio of the germ cells and the X:A ratio of the surrounding soma (Cline and Meyer 1996). The expression of *Sxl*$^+$ in the soma is required in the female germ line. Three genes, *ovo*$^+$, *ovarian tumor*$^+$ and *sans fille*$^+$, are important for growth and differentiation of female, but not male, germ cells. Thus, activation and splicing of *Sex lethal*$^+$ in the ovary is regulated by a different set of proteins from those in the soma (Vincent et al. 2001).

10.6. Are Sex Determination Mechanisms Diverse?

Sex determination mechanisms in insects appear to be diverse (Lauge 1985, Retnakaran and Percy 1985, White 1973, Wrensch and Ebbert 1993, Werren and Beukeboom 1998). Many insects have a **genetic sex determination** system, with genetic differences determining maleness or femaleness. Others appear to have **environmental sex determination**, in which there are no genetic differences between males and females but temperature or host conditions determine the sex. For example, in a few insects the hemolymph of the mother determines the sex of the offspring.

Ploidy levels sometimes are important in sex determination: both sexes of many arthropods are diploid (2n, diplo-diploidy), whereas others have haploid males and diploid females (n and 2n, haplo-diploidy or **arrhenotoky**). Other species consist primarily of diploid females (**thelytoky**), and haploid males rarely are produced. In some species haploid males are produced by the loss of paternally derived chromosomes after fertilization (**parahaploidy**).

Males in apterygote and many pterygote insects are **heterogametic** (males are XO, XY, XXO, XXY, or XYY and females are XX), but in some higher pterygotes (Trichoptera, Lepidoptera) females may be **heterogametic** (ZW).

At least five different models have been proposed to explain sex determination in the haplo-diploid Hymenoptera (Cook 1993, Beukeboom 1995, Dobson and Tanouye 1998a). In the honey bee *Apis mellifera* and the parasitoid *Bracon* (= *Habrobracon*) *hebetor*, sex is

determined by a series of alleles at a single locus (**single locus, multiple-allele model**) (Whiting 1943). In honey bees, the single locus (probably located on chromosome 8) has several alleles (19 so far). Individuals that are heterozygous for this locus are normal fertile (diploid) honey-bee females, hemizygotes (unfertilized haploid eggs) became fertile drones (males), and homozygotes are sterile diploid males with degenerated testes containing reduced quantities of diploid sperm (Beye et al. 1996; 1999). Usually, diploid honey-bee males are eaten by their nestmate workers.

Under the single-locus, multiple-allele model, inbreeding should produce homozygous (diploid) males in the parasitoid *Bracon hebetor*. However, Heimpel et al. (1999) evaluated the diversity of sex-determining alleles in five U.S. populations and estimated that a minimum of 20 alleles were present. This high allelic diversity suggests that the production of diploid males would be rare unless extreme inbreeding were induced.

The haplo-diploid turnip sawfly *Athalia rosae ruficornis*, when inbred, produced both diploid and triploid males, suggesting that sex in this hymenopteran is determined by the single-locus, multiple-allele system (Figure 10.3, Naito and Suzuki 1991). This sex determination system also has been found in the parasitoid *Diadromus pulchellus* (El Agoze et al. 1994), the bee *Apis cerana*, the sawfly *Neodiprion nigroscutum*, the fire ant *Solenopsis invicta*, the stingless bee *Melipona quadrifasciata* (Cook 1993), and the parasitoid *Diadegma chrysostictos* (Butcher et al. 2000).

In other haplo-diploid Hymenoptera, sex is determined by a number of alleles at a series of loci (**multiple-locus, multiple-allele model**). According to this model, females must be

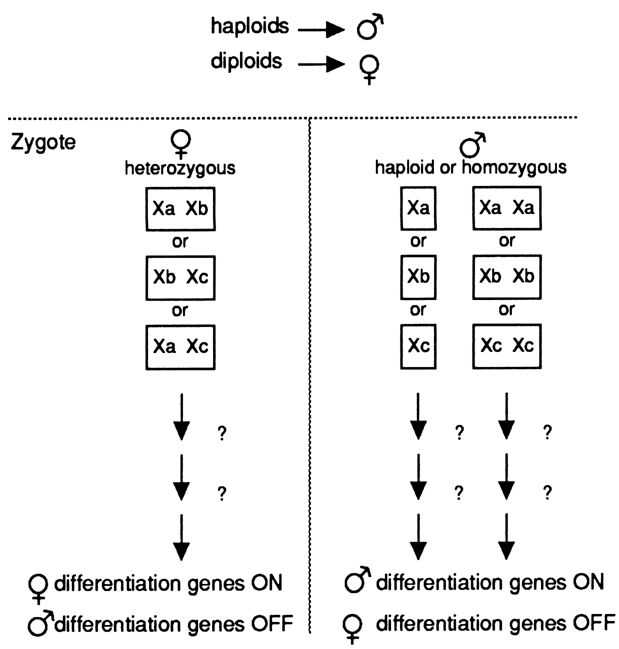


Figure 10.3. Multiple alleles at a single locus determine sex in the haplo-diploid hymenopteran *Athalia rosae*. Under normal conditions, males are haploid and females are diploid, with females heterozygous for the sex-determining locus (X) while males are hemizygous. If individuals become homozygous for an allele of X, perhaps through inbreeding, they become diploid males. (Redrawn from Bownes 1992.)

heterozygous at one or more loci, while haploid males are hemizygous (Crozier 1971). After inbreeding, some diploid individuals are produced, and these are males if they are homozygous for all loci. These two "multiple-allele" models can be combined if the assumption is made that the single-locus model is a special case of the multiple-locus, multiple-allele model. Under this assumption, only one locus has an effect in the first model.

Another model suggests that sex in haplo-diploid (arrhenotokous) Hymenoptera is determined by a balance (**genic balance sex determination model**) between nonadditive male-determining genes and additive female-determining genes scattered throughout the genome (daCunha and Kerr 1957). In this model, maleness genes (m) have noncumulative effects, but femaleness genes (f) are cumulative. Thus, sex is determined by the relationship between f and m. In haploid individuals $m > f$, which results in a male.

10.6.1. Intraspecific Variability

Within a single species, several different sex-determining mechanisms may occur. Many populations of the housefly, *M. domestica*, have five pairs of autosomes and a pair of heterochromatic sex chromosomes; thus, females are XX and males are XY (Figure 10.4). In them, sex is determined by the presence or absence of the Y, which carries a

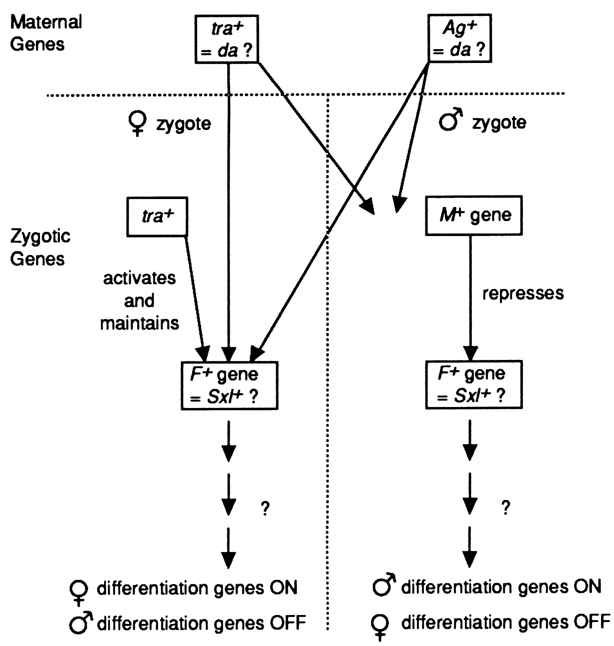


Figure 10.4. Sex determination in the housefly *Musca domestica*. Tra^+ may be equivalent to da^+. F^+ may be equivalent to Sxl^+. tra^+ and Ag^+ gene products are produced by the mother and stored in the egg. In the female zygote these products activate the F^+ gene. The zygote's tra^+ gene must be active to maintain the function of F^+. This leads to expression of female differentiation genes, but the genes lower in the hierarchy are unidentified at present. The M^+ gene product is present in males, which represses the F^+ gene function, so that female differentiation genes are repressed and unknown male differentiation genes are activated. (Modified after Bownes 1992.)

male-determining factor, *M*; the X plays no important role in sex determination. In other strains of *M. domestica*, *both* males and females are XX and have a special autosome that may carry a male-determining factor A^M that determines sex. The A^M component is located on different linkage groups (different chromosomes) in different populations. The presence or absence of the *M* factor seems to be the primary signal for sex determination in these strains. Interestingly, in other strains of *M. domestica*, both males and females have the *M* factors in the homozygous state, and the presence or absence of a female-determining dominant factor (*F*) determines sex. Finally, a dominant maternal-effect mutation, *Arrhenogenous* (*Ag*), has been found in *M. domestica* that causes female progeny to develop into fertile males. A recessive maternal-effect mutation, *transformer*, causes genotypic female progeny carrying no *M* factors to follow the male pathway of sexual development to varying degrees. This suggests that the normal tra^+ gene product is necessary for female determination and/or differentiation and that the gene is expressed during oogenesis and in zygotes (Inoue and Hiroyoshi 1986). Experiments suggest that *M* acts early in embryogenesis to suppress a key gene, perhaps *F*, whose activity is required continuously for development of females, as is Sxl^+ in *Drosophila* (Hilfiker-Kleiner et al. 1993).

In the phorid fly *Megaselia scalaris*, the sex-determining linkage group is not fixed. Different chromosomes serve as the sex-determining pair in different populations. Traut and Willhoeft (1990) estimate that the male-determining factor moved to a different linkage group, thereby creating new Y chromosomes with a frequency of at least 0.06%, which is consistent with the hypothesis that the sex-determining factor is moving by transposition. An alternative explanation is that mutations at multiple sex loci in the genome result in males; however, the high rates of change (0.06%) are higher than expected if due to mutation. The movement of male-determining genes to new chromosomes allows an analysis of sex chromosome evolution (Traut 1994a,b). Analysis of the sex determination cascade in *M. scalaris* indicates that $doublesex^+$ is highly conserved compared to dsx^+ in *D. melanogaster* (Kuhn et al. 2000), but $Sex\text{-}lethal^+$ is not functionally conserved in *M. scalaris* (Sievert et al. 2000). Analyses of other insects also suggest that the base of the sex determination cascade is more highly conserved in function than the upper layer of the cascade (Figure 10.2).

10.6.2. Environmental Effects

Environmental conditions can influence sex determination in some arthropods. Many haplodiploid insects adjust the sex ratio of their progeny based on environmental factors. For example, females of species in the genus *Encarsia* (Hymenoptera: Aphelinidae) develop as **autoparasitoids** of whiteflies (which are considered the primary hosts). Males of the same *Encarsia* species develop as parasitoids of *Encarsia* female pupae, which are considered the secondary hosts. Virgin females deposit unfertilized eggs to produce haploid sons on secondary hosts (females of their own species), but typically do not oviposit in primary hosts (whiteflies), even if they are the only hosts available. When a virgin female does deposit haploid male eggs in a primary host (whiteflies), these eggs usually do not develop, for unknown reasons.

An unusual population of *Encarsia pergandiella* was found in which males could develop on the primary whitefly host. It appears that these haploid males started out as fertilized diploid eggs but became haploid males after the loss of the paternal set of chromosomes shortly after fertilization. This aberrant chromosomal behavior perhaps was caused by a supernumerary chromosome (Hunter et al. 1993).

In the blowfly *Chrysomya rufifacies* (Calliphoridae), females produce *either* female progeny only (**thelygenic** females) *or* male progeny only (**arrhenogenic** females) (Clausen and Ullerich 1990). Thelygenic females are heterozygous for a dominant female-determining maternal effect gene (F'), while arrhenogenic females and males are homozygous for the recessive allele (f). This species lacks differentiated sex chromosomes. DNA sequence homology between the *D. melanogaster* da^+ gene and a polytene band in the genetic sex chromosome pair of *C. rufifacies* was observed by *in situ* hybridization, suggesting that *F* in *C. rufifacies* and da^+ in *D. melanogaster* are equivalent (Clausen and Ullerich 1990).

10.6.3. Postzygotic Sex Determination

In several collembolans, including *Sminthurus viridis* and *Allacma fusca* (suborder Symphpleona), the two sexes differ by having 10 chromosomes in males and 12 in females (Dallai et al. 1999, 2000). Sex determination in these collembolans occurs *after* the zygote forms (rather than at syngamy), when two chromosomes are eliminated in the male embryos in both the somatic and germ-line cells (Dallai et al. 2000). Oogenesis is normal, but spermatogenesis is unusual; daughter cells of the first meiotic division have either six or four chromosomes. The cell receiving four chromosomes degenerates, but the cell with six completes meiosis and produces identical sperm. At fertilization, the pronuclei with six chromosomes fuse to form a zygote with 12 chromosomes. Male embryos then lose two sex chromosomes during the first mitosis, resulting in 10 chromosomes. The mechanism of chromosome elimination during early embryogenesis must be regulated by the genetic constitution of the mother, which means that females could regulate the sex ratio of their progeny. In fact, these species appear to have a female-biased sex ratio. Dallai et al. (2000) suggested that this aberrant meiosis and the large number of females in these species may be considered a step toward the evolution of parthenogenesis.

10.7. A Single Model?

Given the examples of diverse sex-determining systems, is it likely that a single model can describe sex determination in all insects? Nothiger and Steinmann-Zwicky (1985) proposed that all the sex determination mechanisms in insects are variations upon a theme (Figure 10.5). In their model, there is a gene equivalent to Sxl^+, a repressor (R) which inactivates Sxl^+, a gene which activates Sxl^+, and a gene which is equivalent to dsx^+ which is expressed in two alternative forms to interact with one or the other of the two sets of male and female differentiation genes lower in the hierarchy.

Bownes (1992) used their model to compare the sex determination system in the housefly (Figure 10.4). According to the Nothiger and Steinmann-Zwicky (1985) model, the male-determining factor (*M*) in *M. domestica* would correspond to the repressor (R) of Sxl^+. The genes tra^+ and Ag^+ may be equivalent to da^+ in *Drosophila*. The *F* gene of *M. domestica* could be equivalent to Sxl^+. Sxl^+ is involved in dosage compensation in *Drosophila*, but dosage compensation is not needed in species such as *M. domestica* with heterochromatic sex chromosomes (which usually contain few coding regions) or no sex chromosomes. As a result, insects with heterochromatic sex chromosomes or no sex chromosomes can survive mutations of tra^+, Ag^+, and *F*; such mutations can alter sex determination but are not lethal to one sex.

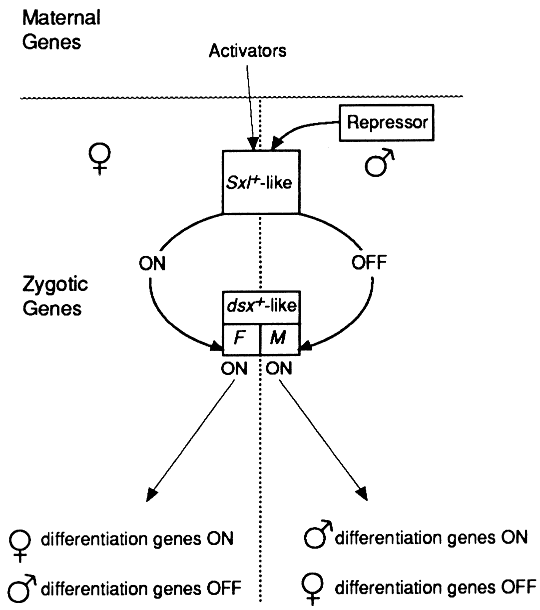


Figure 10.5. Is there a general model for sex determination in insects? This model assumes that activators are produced by the mother that activate an *Sxl*$^+$-like gene in the zygote. Other activators may be produced by the zygote, and the combination results in a functional *Sxl*$^+$-like product in females. Male zygotes produce a repressor of the *Sxl*$^+$-like gene, and no functional product is produced in males. Next, a *dsx*$^+$-like gene is turned on in both females and males, but different products are produced in the two sexes. The different *dsx*$^+$ gene products turn on a subsequent series of genes that result in the differentiation of either males or females. Evidence for this general model is fragmentary at this time. (Redrawn from Bownes 1992.)

Different genera of mosquitoes have several different sex determination systems, but these systems still may conform to the Nothiger and Steinmann-Zwicky model (Bownes 1992). *Anopheles gambiae* and *A. culicifacies* have XY males and XX females. Sex in *Aedes* is determined by a dominant male-determining factor. Intersex flies with phenotypes similar to the *ix*, *dsx*, and *tra* mutants of *Drosophila* have been found in *Aedes aegypti* and *Culex pipiens*. Sex in *Culex* is determined by a single gene on an autosome; *Culex* gynandromorphs have been found, suggesting that sex determination is cell autonomous, as it is in *Drosophila*. The sex of some northern strains of *Aedes* depends upon the temperature at which they are reared, with males transformed into intersexes at higher temperatures. This suggests that an allele equivalent to *ix*$^+$ is temperature sensitive in these populations. In *Culex pipiens*, a sex-linked gene *cercus* (*c*) changes females into intersexes; these intersexes are sterile and fail to take blood meals. It is possible that *cercus*$^+$ is similar to *tra*$^+$, *ix*$^+$, or *dsx*$^+$ of *Drosophila* (Bownes 1992).

Nothiger and Steinmann-Zwicky's (1985) model also might account for sex determination in the haplo-diploid Hymenoptera. According to their model, the multiple alleles which must be heterozygous to confer female development could be mutations of *Sxl*$^+$. These mutations produce an inactive gene product and lead to male development when they are in the homozygous or hemizygous state. However, the different alleles must, in some way, complement each other so that when two different mutations are present in a diploid they are able to make a functional product to produce a female.

Sexual differentiation among different organisms (flies, nematodes, and mammals) has superficially similar patterns of hierarchical control (for reviews, see Marin and Baker 1998, McAllister and McVean 2000). Comparative genetic analysis suggests that the functions of *tra-2*$^+$ and *dsx*$^+$ may be conserved throughout higher eukaryotes. Sex-determining mechanisms are, however, variable, and the function of *Sxl*$^+$ is not conserved among all arthropods.

10.8. Meiotic Drive Can Distort Sex Ratios

Meiotic drive alters the assortment of chromosomes during meiosis so that certain chromosomes are inherited more frequently than expected (>50%). Meiotic drive most frequently is observed when sex chromosome allocation is disrupted (sex chromosome meiotic drive) so that the sex ratio is altered. Whether meiotic drive mechanisms actually modify sex chromosome distribution more frequently is unknown (Lyttle 1993).

Sex chromosome meiotic drive has been found most often in the Diptera, including the *Drosophila obscura*, *melanica*, *tripunctata*, *testacea*, *melanogaster*, and *quinaria* groups, mosquitoes (*Aedes and Culex*), sciarid flies, and stalk-eyed flies (Diopsidae) (Jiggins et al. 1999). The frequency of meiotic drive in other insects is unclear (Jiggins et al. 1999, Jaenike 2001). Three examples of sex chromosome meiotic drive in Diptera are described below, including Segregation Distorter (SD) in *Drosophila*, male drive in the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus*, and meiotic drive in stalk-eyed flies.

10.8.1. Segregation Distorter (SD)

In *Drosophila melanogaster* the SD phenotype is present at low, but stable, frequencies in most field populations. *D. melanogaster* males heterozygous for the (SD) chromosome (SD/SD$^+$) may produce only progeny with the SD chromosome, instead of half with SD and half with SD$^+$, because of the failure of sperm with the SD$^+$ chromosome to mature (Ashburner 1989, Ganetzky 2000). Segregation distortion occurs because the nuclei of the sperm with the normal SD$^+$ chromosome fail to condense normally at sperm maturation. Another gene, the Enhancer locus of SD, *E*(*SD*), is required for the full expression of meiotic drive.

The SD "locus" actually consists of two overlapping genes, one called HS2ST and one called RanGAP. RanGAP is an essential component of a system that transports proteins and RNA molecules into and out of the cell's nucleus. Both HS2ST and RanGAP actually are present twice on the SD chromosome, as opposed to once on SD$^+$ chromosomes, and the tandem duplication is necessary for the segregation distortion (Palopoli et al. 1994). Both genes appear normal in the right hand copy, but the RanGAP gene on the left lacks the last 234 amino acids (Merrill et al. 1999, Ganetzky 2000).

10.8.2. *Distorter* in Mosquitoes

Meiotic drive has been described in the mosquitoes *Aedes aegypti* and *Culex quinquefasciatus*. In both species, a Y-linked gene results in excess males (Wood and Newton 1991). Excess males are produced because X chromosomes are broken during meiosis in males, and thus fewer X than Y chromosomes are transmitted in the sperm, leading to the production of fewer female embryos. The *Distorter* gene (*D*) is linked closely to the sex locus *m/M* and causes the chromosome breakage. Additional genes are involved, and sensitivity to

Distorter is controlled by *m*, the female-determining locus. In some strains, sensitivity is influenced by a second sex-linked gene *t*. Yet another sex-linked gene *A* enhances the effect of *Distorter*. *Distorter* has been found in mosquito populations from Africa, America, Australia, and Sri Lanka, but resistance to it is widespread.

10.8.3. Female-Biased Sex Ratios in Stalk-Eyed Flies

Extreme female-biased sex ratios are found in two sister stalk-eyed fly species, *Cyrtodiopsis dalmanni* and *C. whitei* (Diopsidae), due to a meiotic drive element on the X chromosome (Presgraves et al. 1997). Eye stalks are more exaggerated in males than in females (Wilkinson et al. 1998), and females prefer to mate with males with a long eye span. The long stalks appear to indicate to the female that the male either lacks meiotic drive or can suppress the meiotic drive, thereby increasing the female's fitness by avoiding a biased sex ratio in her progeny (Wilkinson and Reillo 1994). Apparently, there are both autosomal- and Y-linked polymorphisms for resistance to the meiotic drive.

10.8.4. Meiotic Drive as a Pest Management Tool?

Meiotic drive operates as an evolutionary force that can cause an increase in the population frequency of the allele or chromosome which is favored in transmission, even if it confers a disadvantage on its carriers. It has been proposed that meiotic drive might be used to introduce new genes (such as cold-sensitive lethal genes, insecticide-susceptibility genes, or behavior-altering genes that would reduce the negative effects of mosquitoes on humans) into natural populations as a method to achieve control of these important vectors of disease (Wood and Newton 1991). However, much remains to be learned about the stability of meiotic drive mechanisms and the conditions under which they might function in pest management programs.

10.9. Hybrid Sterility

When different species are crossed, hybrid progeny sometimes are produced. However, the progeny may have altered sex ratios, with one sex absent, rare, or sterile. The missing or sterile sex is usually the heterogametic sex (Laurie 1997). This phenomenon is known as **Haldane's Rule**. Hurst and Pomiankowski (1991) suggest that Haldane's Rule only occurs in taxa with sex chromosome-based meiotic drive, such as the Lepidoptera and Diptera. Thus, Haldane's rule may be accounted for in some insects by a loss of suppression of sex ratio distorters when in the novel nuclear cytotype of the hybrid.

Sex ratio distorters that result in unisexual sterility in crosses between different species have been found in many species of *Drosophila*, the dipterans *Musca domestica* and *Glossina morsitans*, the hemipteran *Tetraneura ulmi*, and Lepidoptera (*Acraea encedon*, *Maniola jurtina*, *Danaus chrysippus*, *Philudoria potatoria*, *Mylothris spica*, *Abraxus grossulariata*, *Talaeporia tubulosa*) (Hurst and Pomiankowski 1991).

10.10. *Medea* in *Tribolium*

A new class of selfish genes, *Medea*, was found in the flour beetle *Tribolium castaneum* (Beeman et al. 1992). *Medea* causes a "Maternal-Effect Dominant Embryonic Arrest" that

results in the death of zygotes that do not carry it. If a mother carries *Medea*, any of her offspring who lack this gene die before they pupate. Females who are heterozygous for *Medea* lose half their progeny if they mate with a wild-type male and one-fourth of their progeny when mated to a heterozygous male. It was hypothesized that *Medea* could lead to reproductive isolation and speciation in *T. castaneum*.

A survey of wild populations of *T. castaneum* from Europe, North and South America, Africa and Southeast Asia showed that four different *Medea* alleles were widespread, but absent or rare in Australia and the Indian subcontinent (Beeman and Friesen 1999). Thomson and Beeman (1999) suggest that *Medea* factors are absent from India because a hybrid incompatibility factor (H) is found in the *T. castaneum* populations in India. Apparently H and *Medea* strains of *T. castaneum* are incompatible.

10.11. Cytoplasmic Agents Distort Normal Sex Ratios

A number of cytoplasmically transmitted organisms (bacteria, viruses, protozoans) alter the "normal" sex-determining mechanism(s) in arthropods (Table 10.2). Most are inherited primarily through the oocyte of the mother (cytoplasmically inherited). Cytoplasmic agents that can manipulate their host's sex ratio and promote their own spread are called **cytoplasmic sex ratio distorters**. The spread of a cytoplasmic sex ratio distorter often reduces the fitness of its host and can drive populations to extinction. Sex ratio distorters are usually suspected if crosses produce a heavily female-biased sex ratio, although meiotic drive and hybrid dysgenesis agents are other possible mechanisms (Hurst 1993).

Ebbert (1991, 1993) described at least 50 cases in which cytoplasmically inherited organisms alter sex ratios in the Diptera, Heteroptera, Coleoptera, Lepidoptera, and Acari (mites). Such sex ratio distorters may be widespread, but undetected, in other arthropods because sex ratios rarely are assessed by making single-pair crosses. Transmission rates of these agents typically are high, although a few progeny may fail to become infected. The altered sex-ratio conditions are found in natural populations at frequencies ranging from "low to high." The infections may reduce fitness of the hosts and reduce egg hatch or larval survival in the progeny of infected females.

10.11.1. Spiroplasmas

The sex ratio condition of *Drosophila willistoni*, and related neotropical *willistoni* group species, is due to a spiroplasma (Ashburner 1989, Williamson et al. 1999). The spiroplasmas are maternally inherited, transovarially transmitted, and lethal to male embryos. Spiroplasmas can be transmitted between species by injecting hemolymph, but spiroplasmas from different species are different, and a different virus is associated with each spiroplasma. When spiroplasmas from different species are mixed, they clump because the viruses lyse the spiroplasmas of the other species.

10.11.2. L-Form Bacteria

The *Drosophila paulistorum* complex contains six semispecies (subgroups derived from a single species that are thought to be in the process of speciation) that do not normally interbreed. When they are crossed in the laboratory, fertile daughters and sterile sons are

produced. Streptococcal L-forms (bacteria) that are associated with the sterility have been isolated and cultured in artificial media (Somerson et al. 1984). The L-forms are transferred through the egg cytoplasm, and each semispecies appears to have a different microorganism. The L-forms can be microinjected into females and can produce the expected male sterility. This suggests that an L-form normally has a benign relationship with its own host; however, if it is transferred to a closely related host, sterility is induced.

10.11.3. *Wolbachia*

These rickettsial-like bacteria are one of the most commonly described cytoplasmically inherited microorganisms in arthropods. *Wolbachia* are gram-negative rods that cannot be cultured easily outside their hosts. They are widespread: 17 to 76% of all arthropod species contain them (Werren et al. 1995, Werren 1997, Jeyaprakash and Hoy 2000). Knowledge of the evolution and physiological and phenotypic effects of *Wolbachia* on most of their hosts remains limited (Rigaud 1999). One of their effects in arthropods is to alter sex ratio (Rigaud and Rousset 1996, O'Neill et al. 1997, Bourtzis and O'Neill 1998, Cook and Butcher 1999, Stouthamer et al. 1999, Vavre et al. 2000, Stevens et al. 2001). An overview of *Wolbachia* as endosymbionts was provided in Chapter 4, so this discussion will focus on the effects of *Wolbachia* on sex determination and sex ratio.

Wolbachia infect isopods (Crustacea), including *Armadillidium album*, *Ligia oceanica*, *A. nasatum*, *Porcellionides pruinosus*, *Chaetophiloscia elongata*, and *Spaeroma rugicauda* (Rigaud and Rousset 1996). Some *Wolbachia*-infected isopods produce female-biased broods because the *Wolbachia* change genetic males (homogametic ZZ individuals) into functional "females." These ZZ individuals are chromosomally male, but phenotypically appear *and function* as females. "Daughters" of infected mothers produce all-female or highly female-biased progeny, resulting in isopod lineages that are chromosomally males (ZZ) but are functional females (Rigaud and Rousset 1996). Interestingly, there has been speculation that some *Wolbachia* genes have been transferred to the isopod nuclear genome—reminiscent of the movement of genes over evolutionary time from the mitochondrion (originally a microbial symbiont) to the nuclear genome of eukaryotes. It will be interesting to learn if any *Wolbachia* genes have been transferred to the nuclear genome of insects.

Wolbachia can cause thelytoky, male killing, cytoplasmic incompatibility, and female mortality in arthropods (Bandi et al. 1999, Rousset and Raymond 1991, Stouthamer 1997, Majerus et al. 1999, Vavre et al. 2000). Some *Wolbachia* improve fertility or vigor while others appear to decrease these traits in their hosts. Some species appear to have *Wolbachia* only in their germ line (ovaries and testes), whereas others have *Wolbachia* in somatic tissues as well (Dobson et al. 1999). Large numbers of *Wolbachia* have been found in ovaries and testes of populations with cytoplasmic incompatibilities. Cytoplasmic incompatibility occurs when *Wolbachia*-infected males mate with uninfected females, resulting in a failure to produce progeny in diplo-diploid species. The reciprocal cross should produce a normal number of progeny (described in Chapter 4).

Wolbachia may case thelytoky in the Hymenoptera, which typically are arrhenotokous (Stouthamer 1997). *Wolbachia*-induced thelytoky (parthenogenesis in which only females are known) has been found in the Tenthredinoidea, Signiforidae, and Cynipoidea (Stouthamer 1997), as well as at least 70 species of parasitoids (Aphelinidae, Encyrtidae, Eulophidae, Pteromalidae, Torymidae, Trichogrammatidae, Cynipidae, Eucoilidae, Braconidae, Ichneumonidae, Proctotrupoidae) (Stouthamer 1997, Cook and Butcher 1999). A number of hymenopteran parasitoid species have arrhenotokous and thelytokous strains.

In thelytokous populations of parasitoids, unfertilized eggs give rise to females. A number of thelytokous parasitoids (*Ooencyrtus submetallicus*, *Pauridia peregrina*, *Trichogramma* sp., *Ooencyrtus fecundus*) produce a few males, usually less than 5%, when reared at temperatures over 30°C (Stouthamer 1997), suggesting incomplete transmission of *Wolbachia* or a low titer of *Wolbachia*. Sometimes, these rare males have been shown to mate and transfer sperm to conspecific females, indicating that the males retained normal vigor and fertility. In other cases, the rare males were infertile, suggesting that the *Wolbachia* infection had existed for a long time in the population, which could have relaxed selection for essential fertility genes over evolutionary time. In addition to heat, several antibiotics (tetracycline hydrochloride, sulfamethoxazole, and rifampin) can induce the production of males in some thelytokous parasitoid populations.

The cytogenetic changes that occur during meiosis to restore an unfertilized haploid egg to diploidy (thus permitting thelytoky) have been studied in *Trichogramma*. In the eggs of *Wolbachia*-infected *Trichogramma* females, meiosis progresses to the stage of a single haploid pronucleus and the diploid chromosome number is restored during the first mitotic division. Thus, during anaphase, the two identical sets of chromosomes do not separate, and the result is a single nucleus containing two copies of the same set of chromosomes, resulting in a female that is completely homozygous at all loci (Stouthamer 1997).

How might parasitoids (or other insects) be infected with *Wolbachia*? Phylogenetic analysis of *Wolbachia* suggests that both horizontal and vertical transfer of *Wolbachia* occurs among insects (Jeyaprakash and Hoy 2000). Horizontal transmission of thelytoky-inducing *Wolbachia* from one parasitoid strain to another within a shared lepidopteran host was demonstrated by Huigens et al. (2000). Offspring of uninfected *Trichogramma* females can acquire sufficient thelytoky-inducing *Wolbachia* to express the trait when they share a host egg with progeny of *Wolbachia*-infected females. The process by which the uninfected *Trichogramma* larvae acquire the *Wolbachia* remains unclear. However, this intraspecific horizontal transfer suggests that interspecific horizontal transfers from parasitoid to parasitoid could occur when the parasitoids share a common host.

Wolbachia infections causing thelytoky are hypothesized to be a mechanism that contributes to the process of speciation. For example, some populations of the parasitoid *Encarsia formosa* no longer have males, so that these populations essentially become clonal and over time could differentiate genetically. *Wolbachia*-induced incompatibility is thought to precede hybrid incompatibilities in the parasitoid *Nasonia* (Bordenstein et al. 2001). Under this scenario, an uninfected ancestral population gives rise to two geographically isolated daughter populations. If each population is infected with a different strain of *Wolbachia*, the populations could become reproductively isolated as a result of their infections. The role of *Wolbachia* in speciation is controversial and, according to some, unproven (Werren 1997, Hurst and Schilthuizen 1998, Shoemaker et al. 1999, Rokas 2000).

Wolbachia can influence mating behavior and kill males in populations of the butterfly *Acraea encedon* across Africa. In many populations, females produce only female progeny, while other populations produce both males and females in a normal 1:1 sex ratio. The production of all-female progeny is caused by a *Wolbachia* that kills males (Jiggins et al. 1998). *A. encedon* typically deposit clutches of 50 to 300 eggs, and newly hatched larvae often cannibalize unhatched eggs, only gradually dispersing into smaller groups. Jiggins et al. (1998) speculated that the evolution of male-killing by *Wolbachia* may be favored when the behavior and ecology of a species makes antagonistic interactions between siblings or sib cannibalism likely. Under field conditions, *Wolbachia* infections in *A. encedon* females may result in populations with a serious shortage of males. As a consequence, the mating behavior

of *Wolbachia*-infected *A. encedon* has been altered. Normally, males seek out and compete for individual females near larval food plants. However, when male-killing *Wolbachia* are present in high frequency in a population, females instead form dense aggregations in grassy areas near trees, perhaps to attract rare males as mates.

Wolbachia have been proposed as "drive" mechanisms to introduce transgenes into arthropod populations (Turelli and Hoffman 1999).

10.12. A Mite Consisting Only of Haploid Females

The false spider mite *Brevipalpus phoenicis* (Acari: Tenuipalpidae) has been found to consist entirely of females that have only a haploid chromosome set (Weeks et al. 2001). This unusual genetic system is due to the presence of an endosymbiotic bacterium (not *Wolbachia*) which has feminized haploid males. An analysis of 16S ribosomal DNA sequences suggests the bacterium's closest relative is an undescribed endosymbiont found in the tick *Ixodes scapularis*. If female mites are treated with antibiotics, about half their progeny become male. How the bacterium induces feminization of genetic males is unknown.

10.13. Paternal Sex Ratio Chromosomes and Cytoplasmic Incompatibility in *Nasonia*

Sex ratio in the parasitoid *Nasonia vitripennis* can be altered by at least two different mechanisms. Some natural populations of *N. vitripennis* carry a supernumerary or **B chromosome** that causes a condition called **paternal sex ratio (PSR)**. B chromosomes are found in many plant and animal species and are small nonvital chromosomes consisting mostly of heterochromatin. B chromosomes have few genes and often cause a small fitness cost to their host, making them "selfish" genetic elements. Some B chromosomes are thought to be derived from normal chromosomes and may be transmitted at higher rates than expected, exhibiting "drive" (Jones and Rees 1982).

The PSR chromosome is carried only by male *N. vitripennis* and is transmitted via sperm to fertilized eggs. After an egg is fertilized by a PSR-bearing sperm, the paternally derived chromosomes condense into a chromatin mass and are lost, leaving only the maternal chromosomes. The PSR chromosome itself survives, changing fertilized diploid (female) eggs into haploid PSR males. PSR is unusual in its ability to destroy the genome of its carrier each generation (Werren et al. 1987, Nur et al. 1988, Beukeboom and Werren 1992, 1993, Beukeboom et al. 1992, Reed and Werren 1995).

Where did the PSR chromosome come from? The PSR chromosome has sequences that are homologous with autosomal sequences of *Nasonia giraulti*, *N. longicornis*, and *Trichomalopsis dubius*, but not with *N. vitripennis* (Eickbush et al. 1992). The PSR chromosome could have been present prior to the divergence of the genera *Trichomalopsis* and *Nasonia* (Eickbush et al. 1992). Alternatively, PSR may have crossed the species barrier more recently (horizontal transfer) through a series of interspecific transfers between species capable of mating (Dobson and Tanouye 1998b). Experimental interspecific transfer of the PSR was successful after these species were cured of *Wolbachia*, which causes cytoplasmic incompatibility between them (Dobson and Tanouye 1998b). The transferred PSR chromosome continued to function in both recipient species.

The sex ratio of *Nasonia vitripennis* is modified by other non-Mendelian factors, including “Son-killer,” a maternally transmitted bacterium that prevents development of unfertilized male eggs, and “Maternal Sex Ratio,” a cytoplasmically inherited agent that causes female wasps to produce nearly 100% daughters.

10.14. Male Killing in Coccinellidae

Male killing is associated with a variety of microorganisms. The Coccinellidae appear particularly prone to invasion by male-killing endosymbionts, with four different groups (*Rickettsia*, *Spiroplasma*, Flavobacteria, and *Wolbachia*) identified (Majerus and Hurst 1997, Majerus et al. 1999, Hurst and Jiggins 2000). Coccinellids may be especially susceptible to invasion by and establishment of male-killing microbes because of their biology. Coccinellids feed on aphids and lay eggs in tight batches, which promotes sibling egg cannibalism and significant levels of mortality of newly hatched larvae due to starvation (Majerus et al. 1999).

The evolution of male killing may have evolved because the bacteria are almost exclusively transmitted vertically from mother to eggs. As a result, bacteria in male hosts are at an evolutionary dead end, so male-killing has a fitness cost of zero from the bacterial point of view (Randerson et al. 2000). Furthermore, the death of male embryos could augment the fitness of the remaining female brood by providing food to those females carrying the clonal relatives of the male-killing bacteria (Randerson et al. 2000).

10.15. Sex and the Sorted Insects

Resolving the molecular genetics of sex determination in arthropods and learning how to modify sex ratio or fertility will have both theoretical and applied applications and could lead to improved genetic control of pests or useful genetic modifications of beneficial biological control agents.

10.15.1. Genetic Control

Genetic control of pest insects represents an attractive alternative to chemical control in terms of safety, specificity, and the limited negative impact it has upon the environment. The screwworm (*Cochliomyia hominivorax*) eradication campaign demonstrates what can be achieved with mass releases of males sterilized by irradiation (Box 10.1). The principle of sterile insect releases has been applied to other pest insect species, including the Mediterranean fruit fly (*Ceratitis capitata*), tsetse flies (*Glossina palpalis* and *G. morsitans*), mosquitoes (*Anopheles albimanus*), codling moth (*Cydia pomonella*), and ticks (LaChance 1979).

Sterile insect release programs usually require only males, but both sexes must be reared. Not only is it expensive to rear large numbers of “useless” females, but, in the case of species that vector disease or annoy or bite humans or domestic animals, it is undesirable to release any females, sterile or not! As a result, various genetic methods have been used to develop “genetic sexing strains,” strains that make it easy to separate males and females. For example, slight differences in size or color of pupae have been used to sort out the undesirable females during mass rearing. Most genetic sexing strains are based on

Box 10.1. Eradication of the screwworm from North America

The genetic control method used to eradicate the screwworm *Cochliomyia hominivorex* is called the "sterile insect release method" (SIRM) or "sterile insect technique" (SIT). The SIRM involves mass rearing and sterilization of males by chemicals or irradiation, and their subsequent release to mate with wild females. Because females of the screwworm mate only once, any wild female mating with a sterile male fails to contribute progeny to the next generation (Knipling 1955). When an excess of sterile males (compared to the number of wild males) is released, populations decline in a predictable manner, ultimately becoming extinct. Because absolute population densities of *C. hominivorex* were often low in the United States, the number of sterile males that had to be released could be produced in "fly factories."

The screwworm eradication program was initiated in Florida with small-scale trials on Sanibel Island in 1951. The results were promising, and the project was geared up to cover the state of Florida and then the southeastern United States. The screwworm was declared eradicated from the southeastern United States in 1959, one year ahead of schedule. Eradication was achieved in a surprisingly short time because of the combined effects of a severe winter in Florida during 1957–1958, which greatly reduced the overwintering screwworm population, and a 17-month eradication program beginning in July 1958 that cost approximately $7 million and involved the release of almost 9 billion sterile screwworm flies over an area of approximately 56,000 square miles (Meadows 1985).

Since 1959, the livestock industries of Florida and adjacent states have saved at least $20 million each year because the screwworm is no longer present; actual benefits are even greater in today's dollars (Meadows 1985). Furthermore, the elimination of losses due to the deaths of livestock and the elimination of labor and control costs are only part of the benefits; loss of wildlife to screwworm attack also was eliminated.

The success of the SIRM program in the southeastern United States led the cattle growers of Texas to mount, in collaboration with the state and the U.S. Department of Agriculture, a similar but much more ambitious program in the southwestern United States in the 1960s (Bushland 1985). This program required more time and effort because the area in which the screwworm was to be eradicated bordered on a front 2400 kilometers long, stretching from the Gulf of Mexico to the Pacific Ocean. Despite this challenge, and some setbacks with quality control and reinvasion of flies from Mexico, both Texas and New Mexico were declared "screwworm free" in 1964.

The SIRM program was moved into Arizona and California in 1965, and by 1966 the entire United States could be declared free of screwworms. To reduce the likelihood that the screwworm would reinvade the United States from Mexico, the program was expanded into Mexico in 1972, with the goal of eradicating the screwworm all the way south to the Isthmus of Tehuantepec (Pineda-Vargas 1985).

After successfully eliminating the pest in Mexico, the SIRM program was expanded to cover all of Central America (Snow et al. 1985, Wyss 2000). Screwworms were

Box 10.1. continued

eliminated from Guatemala between 1988 and 1994, from Belize between 1988 and 1994, from El Salvador between 1991 and 1995, from Honduras between 1991 and 1996, from Nicaragua between 1992 and 1998, from Costa Rica between 1995 and 1999, and from Panama between 1997 and 2000. These eradication programs were carried out so that a barrier zone could be set up at the 90-km-wide Isthmus of Panama, which is easier to maintain compared to the 2400-km border that the United States and Mexico share. This barrier zone is being maintained by a combination of quarantines and mass releases of sterile screwworms.

Benefits of this massive, and expensive, screwworm eradication program are great (Wyss 2000). In 1996, the producer benefits in the United States, Mexico, and Central America were estimated to be $796 million, $292 million, and $77.9 million annually, respectively. These benefits were due to decreases in deaths of livestock, reduced veterinary services, medicines, insecticides, inspections and handling costs, as well as increases in meat and milk production. The estimated benefit-to-cost ratios for the eradication programs average 12.2:1 for Central America to 18:1 for the United States and Mexico (Wyss 2000).

maintaining marker genes (such as *white pupa* or a temperature-sensitive lethal) within translocations. However, because translocations can undergo recombination in the region between the translocation breakpoint and the marker gene, the strains are not completely stable. As a result, if no practical means exist to remove the recombinants, an increasing number of undesirable females will be reared and released.

Ideally, a genetic sexing method would produce only males of high quality and vigor to compete with wild males for female mates. Because an all-male colony will be difficult to maintain (!), this character ideally would be a conditional trait, perhaps dependent upon temperature or some other environmental cue. Developing genetic sexing systems based on transgenic methods could result in more stable lines than those based on translocations. Likewise, developing transgenic methods to cause male sterility also could be beneficial to an SIRM program. Sterilizing males by irradiation makes them less fit because it causes general somatic damage. Eliminating this fitness loss could allow fewer males to be released, also resulting in a significant savings in program costs.

An unusual approach to obtaining sterile males for genetic control programs involves producing sterile backcross males (Makela and Huettel 1979). Viable hybrid progeny are produced when *Heliothis virescens* male and *H. subflexa* female moths mate; all F_1 females are fertile when backcrossed to males from either parental species. However, most of the F_1 males are sterile, and if hybrid females are backcrossed through successive generations to *H. virescens* males, all male progeny will be sterile in later generations. Maternal inheritance is involved in this hybrid male sterility, but the precise mechanism remains controversial. One potential mechanism is based on incompatibility between sperm mitochondria in backcross males and the cytoplasm, which is derived from the mothers (Miller et al. 1986). An alternative explanation is that incompatibility is caused by interactions between maternally inherited microorganisms and the paternal genetic material in the nucleus. Evidence for this hypothesis is based on the analysis by the PCR of 16S rRNA sequences from microorganisms associated with *H. virescens* and *H. subflexa* (Krueger et al. 1993).

10.15.2. Genetic Improvement of Parasitoids

Genetic improvement of parasitoids reared for augmentative biological control could be achieved if the proportion of females produced in the rearing program could be increased. Experiments have been conducted to determine if it is possible to artificially transfer thelytoky-inducing strains of *Wolbachia* from one *Trichogramma* species to another so that the new strain could be improved by increasing the proportion of female progeny (Grenier et al. 1998). *Wolbachia* isolated from *T. pretiosum* was transferred by microinjection into the pupae of an uninfected species, *T. dendrolimi*. The *Wolbachia* were found in the recipient species 26 generations after the transfer, but only a partial level of thelytoky was observed, perhaps because the density of symbionts was too low or because symbiont–host interactions interfered with the expression of the thelytoky phenotype.

References Cited

Antolin, M. F. (1999). A genetic perspective on mating systems and sex ratios of parasitoid wasps. *Res. Popul. Ecol.* **41**: 29–37.

Ashburner, M. (1989). Drosophila. *A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Baker, B. S. (1989). Sex in flies: the splice of life. *Nature* **340**: 521–524.

Baker, B. S., M. Gorman, and I. Marin (1994). Dosage compensation in *Drosophila*. *Annu. Rev. Genet.* **28**: 491–521.

Bandi, C., B. Slatko, and S. L. O'Neill (1999). *Wolbachia* genomes and the many faces of symbiosis. *Parasitol. Today* **15**: 428–429.

Barton, N. H., and B. Charlesworth (1998). Why sex and recombination? *Science* **281**: 1986–1990.

Beeman, R. W., and K. S. Friesen (1999). Properties and natural occurrence of maternal-effect selfish genes ("Medea" factors) in the red flour beetle, *Tribolium castaneum*. *Heredity* **82**: 529–534.

Beeman, R. W., K. S. Friesen, and R. E. Denell (1992). Maternal-effect selfish genes in flour beetles. *Science* **256**: 89–92.

Belote, J. M., M. B. McKeown, D. J. Andrew, T. N. Scott, M. F. Wolfner, and B. S. Baker (1985), pp. 605–614. In: *Molecular Biology of Development*. Vol. 50. Cold Spring Harbor Symp. Quant. Biol. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Beukeboom, L. W. (1995). Sex determination in Hymenoptera: a need for genetic and molecular studies. *BioEssays* **17**: 813–817.

Beukeboom, L. W., and J. H. Werren (1992). Population genetics of a parasitic chromosome: experimental analysis of PSR in subdivided populations. *Evolution* **46**: 1257–1268.

Beukeboom, L. W., and J. H. Werren (1993). Deletion analysis of the selfish B chromosome, paternal sex ratio (PSR), in the parasitic wasp *Nasonia vitripennis*. *Genetics* **133**: 637–648.

Beukeboom, L. W., K. M. Reed, and J. H. Werren (1992). Effects of deletions on mitotic stability of the paternal-sex-ratio (PSR) chromosome from *Nasonia*. *Chromosoma* **102**: 20–26.

Beye, M., R. F. A. Moritz, R. H. Crozier, and Y. C. Crozier (1996). Mapping the sex locus of the honeybee (*Apis mellifera*). *Naturwissenschaften* **83**: 424–426.

Beye, M., G. J. Hunt, R. E. Page, M. K. Fondrk, L. Grohmann, and R. F. A. Moritz (1999). Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics* **153**: 1701–1708.

Bordenstein, S. R., F. P. O'Hara and J. H. Werren (2001). *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature* **409**: 707–710.

Bourtzis, K., and S. O'Neill (1998). *Wolbachia* infections and arthropod reproduction. *BioScience* **48**: 287–293.

Bownes, M. (1992). Molecular aspects of sex determination in insects, pp. 76–100. In: *Insect Molecular Science*. J. M. Crampton and P. Eggleston, Eds. Academic Press, London.

Bull, J. J. (1983). *Evolution of Sex Determining Mechanisms*. Benjamin/Cummings, London.

Bushland, R. C. (1985). Eradication program in the southwestern United States, pp. 12–15. In: *Symposium on Eradication of the Screwworm from the United States and Mexico.* O. H. Graham, Ed. Misc. Publ. (62). Entomol. Soc. Amer., College Park, MD.

Butcher, R. D. J., W. G. F. Whitfield, and S. F. Hubbard (2000). Single-locus complementary sex determination in *Diadegma crysostictos* (Gmelin) (Hymenoptera: Ichneumonidae). *Heredity* **91**: 104–111.

Charnov, E. (1982). *The Theory of Sex Allocation.* Princeton University Press, Princeton, NJ.

Clausen, S., and F. H. Ullerich (1990). Sequence homology between a polytene band in the genetic sex chromosomes of *Chrysomya rufifacies* and the *daughterless* gene of *Drosophila melanogaster*. *Naturwissenschaften* **77**: 137–138.

Cline, T. W., and B. J. Meyer (1996). Vive la difference: males vs females in flies vs worms. *Annu. Rev. Genet.* **30**: 637–702.

Cook, J. M. (1993). Sex determination in the Hymenoptera: a review of models and evidence. *Heredity* **71**: 421–435.

Cook, J. M., and R. D. J. Butcher (1999). The transmission and effects of *Wolbachia* bacteria in parasitoids. *Res. Popul. Ecol.* **41**: 15–28.

Crow, J. F. (1994). Advantages of sexual reproduction. *Dev. Genet.* **15**: 205–213.

Crozier, R. H. (1971). Heterozygosity and sex determination in haplo-diploidy. *Am. Naturalist* **105**: 339–412.

Dallai, R., P. P. Fanciulli, and F. Frati (1999). Chromosome elimination and sex determination in springtails (Insecta, Collembola). *J. Exp. Zool.* **285**: 215–225.

Dallai, R., P. P. Fanciulli, and F. Frati (2000). Aberrant spermatogenesis and the peculiar mechanism of sex determination in Symphypleonan Collembola (Insecta). *Heredity* **91**: 351–358.

daCunha, A. D., and W. E. Kerr (1957). A genetical theory to explain sex determination by arrhenotokous parthenogenesis. *Forma et Functio* **1**(**4**): 33–36.

DeLoof, A., and R. Huybrechts (1998). "Insects do not have sex hormones": A myth? *Gen. Comp. Endocrinol.* **111**: 245–260.

Dobson, S. L., and M. A. Tanouye. 1998a. Evidence for a genomic imprinting sex determination mechanism in *Nasonia vitripennis* (Hymenoptera: Chalcidoidea). *Genetics* **149**: 233–242.

Dobson, S. L., and M. A. Tanouye. 1998b. Interspecific movement of the paternal sex ratio chromosome. *Heredity* **81**: 261–269.

Dobson, S. L., K. Bourtzis, H. R. Braig, B. F. Jones, W. Zhou, F. Rousset and S. C. O'Neill (1999). *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* **29**: 153–160.

Ebbert, M. A. (1991). The interaction phenotype in the *Drosophila willistoni*–spiroplasma symbiosis. *Evolution* **45**: 971–988.

Ebbert, M. A. (1993). Endosymbiotic sex ratio distorters in insects and mites, pp. 150–191. In: *Evolution and Diversity of Sex Ratio in Insects and Mites*. D. L. Wrensch, and M. A. Ebbert, Eds. Chapman, and Hall, New York.

Eickbush, D. G., T. H. Eickbush, and J. H. Werren (1992). Molecular characterization of repetitive DNA sequences from a B chromosome. *Chromosoma* **101**: 575–583.

El Agoze, M., J. M. Drezen, S. Renault, and G. Periquet (1994). Analysis of the reproductive potential of diploid males in the wasp *Diadromus pulchellus* (Hymenoptera: Ichneumonidae). *Bull. Entomol. Res.* **84**: 213–218.

Ephrussi, A., and R. Lehmann (1992). Induction of germ cell formation by *oskar*. *Nature* **358**: 387–392.

Fialho, R. E., and L. Stevens (2000). Male-killing *Wolbachia* in a flour beetle. *Proc. R. Soc. London B* **267**: 1469–1474.

Ganetzky, B. (2000). Tracking down a cheating gene. *Am. Sci.* **88**: 128–135.

Gherna, R. L., J. H. Werren, W. Weisburg, R. Cote, C. R. Woese, L. Mandelco and D. J. Brenner (1991). *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *Intern. J. System. Bacteriol.* **41**: 563–565.

Greenspan, R. J., and J. F. Ferveur (2000). Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**: 205–232.

Grenier, S., B. Pintureau, A. Heddi, F. Lassabliere, C. Jager, C. Louis and C. Khatchadourian (1998). Successful horizontal transfer of *Wolbachia* symbionts between *Trichogramma* wasps. *Proc. R. Soc. London B* **265**: 1441–1445.

Hackstein, J. H. P., and R. Hochstenbach (1995). The elusive fertility genes of *Drosophila*: the ultimate haven for selfish genetic elements. *Trends Genet.* **11**: 195–200.

Hamilton, W. D. (1967). Extraordinary sex ratios. *Science* **156**: 477–488.

Hamilton, W. D., R. Axelrod, and R. Tanese (1990). Sexual reproduction as an adaptation to resist parasites (a review). *Proc. Natl. Acad. Sci. USA* **87**: 3566–3573.

Hartl, D. L., and S. W. Brown (1970). The origin of male haploid genetic systems and their expected sex ratio. *Theor. Pop. Biol.* **1**: 165–190.

Heimpel, G. E., M. F. Antolin, and M. R. Strand (1999). Diversity of sex-determining alleles in *Bracon hebetor*. *Heredity* **82**: 282–291.

Heinrich, J. C., and M. J. Scott (2000). A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program. *Proc. Natl. Acad. Sci. USA* **97**: 8229–8232.

Hennig, W. (1993). Conventional protein coding genes in the *Drosophila* Y chromosome: Is the puzzle of the fertility gene function solved? *Proc. Natl. Acad. Sci. USA* **90**: 10904–10906.

Hilfiker-Kleiner, D., A. Dubendorfer, A. Hilfiker, and R. Nothiger. 1993. Developmental analysis of two sex-determining genes, *M* and *F*, in the housefly, *Musca domestica*. *Genetics* **134**: 1187–1194.

Hochstenbach, R., R. Brand, and W. Hennig (1994). Transcription of repetitive DNA sequences in the lampbrush loop pair *Nooses* formed by sterile alleles of fertility gene *Q* on the Y chromosome of *D. hydei*. *Mol. Gen. Genet.* **244**: 653–660.

Huigens, M. E., R. F. Luck, R. H. G. Klaassen, M. F. P. M. Maas, M. J. T. N. Timmermans, and R. Stouthamer (2000). Infectious parthenogenesis. *Nature* **405**: 178–179.

Hunter, M. S., U. Nur, and J. H. Werren (1993). Origin of males by genome loss in an autoparasitoid wasp. *Heredity* **70**: 162–171.

Hurst, L. D. (1993). The incidences, mechanisms and evolution of cytoplasmic sex ratio distorters in animals. *Biol. Rev.* **68**: 121–193.

Hurst, G. D. D., and F. M. Jiggins (2000). Male-killing bacteria in insects: mechanisms, incidence and implications. *Emerging Infect. Dis.* **6**: 329–336.

Hurst, L. D., and A. Pomiankowski (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* **128**: 841–858.

Hurst, G. D. D., and M. Schilthuizen (1998). Selfish genetic elements and speciation. *Heredity* **80**: 2–8.

Hurst, G. D. D., T. M. Hammarton, C. Bandi, T. M. O. Majerus, D. Bertrand and M. E. N. Majerus (1997). The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria. *Genet. Res.* **70**: 1–6.

Hurst, G. D. D., J. H. G. V. D. Schulenburg, T. M. O. Majerus, D. Bertrand, I. A. Zakarov, J. Baungaard, W. Volkl, R. Stouthamer and M. E. N. Majerus (1999a). Invasion of one insect species, *Adalia bipunctata*, by two different male-killing bacteria. *Insect Mol. Biol.* **8**: 133–139.

Hurst, G. D. D., F. M. Jiggins, J. H. G. V. D. Schulenburg, D. Bertrand, S. A. West, I. I. Goriacheva, I. A. Zakharov, J. H. Werren, R. Stouthamer, and M. E. N. Majerus (1999b). Male-killing *Wolbachia* in two species of insect. *Proc. R. Soc. London B* **266**: 735–740.

Inoue, H., and T. Hiroyoshi (1986). A maternal-effect sex-transformation mutant of the housefly. *Musca domestica L. Genetics* **112**: 469–482.

Jaenike, J. (2001). Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32**: 2549.

Janzer, B., and M. Steinmann-Zwicky (2001). Cell-autonomous and somatic signals control sex-specific gene expression in XY germ cells of *Drosophila*. *Mech. Dev.* **100**: 3–13.

Jeyaprakash, A., and M. A. Hoy (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of 63 arthropod species. *Insect Mol. Biol.* **9**: 393–405.

Jiggins, F. M., G. D D. Hurst, and M. E. N. Majerus (1998). Sex ratio distortion in *Acraea encedon* (Lepidoptera: Nymphalidae) is caused by a male-killing bacterium. *Heredity* **81**: 87–91.

Jiggins, F. M., G. D. D. Hurst, and M. E. N. Majerus (1999). How common are meiotically driving sex chromosomes in insects? *Am. Naturalist* **154**: 481–483.

Johanowicz, D. J., and M. A. Hoy (1996). *Wolbachia* in a predator–prey system: 16S ribosomal DNA analysis of two phytoseiids (Acari: Phytoseiidae) and their prey (Acari: Tetranychidae). *Ann. Entomol. Soc. Amer.* **89**: 435–441.

Johanowicz, D. L., and M. A. Hoy (1998). Experimental induction and termination of non-reciprocal reproductive incompatibilities in a parahaploid mite. *Entomol. Exp. Appl.* **87**: 51–58.

Jones, R. N., and H. Rees (1982). *B Chromosomes*. Academic Press, New York.

Keightley, P. D., and A. Eyre-Walker (2000). Deleterious mutations and the evolution of sex. *Science* **290**: 331–333.

Kelley, R. L., and M. I. Kuroda (2000). Noncoding RNA genes in dosage compensation and imprinting. *Cell* **103**: 9–12.

Knipling, E. F. (1955). Possibilities of insect control or eradication through the use of sexually sterile males. *J. Econ. Entomol.* **48**: 459–462.

Knipling, E. F. (1985). Sterile insect technique as a screwworm control measure: the concept and its development, pp. 4–8. In: *Symposium on Eradication of the Screwworm from the United States and Mexico*. O. H. Graham, Ed. Misc. Publ. (62). Entomol. Soc. Amer., College Park, MD.

Kopp, A., I. Duncan, and S. B. Carroll (2000). Genetic control and evolution of sexually dimorphic characters. *Nature* **408**: 553–559.

Krueger, C. M., M. E. Degrugillier, and S. K. Narang (1993). Size difference among 16S rRNA genes from endosymbiotic bacteria found in testes of *Heliothis virescens, H. subflexa* (Lepidoptera: Noctuidae) and backcross sterile male moths. *Fla. Entomol.* **76**: 382–390.

Kuhn, S., V. Sievert, and W. Traut (2000). The sex-determining gene *doublesex* in the fly *Megaselia scalaris*: conserved structure and sex-specific splicing. *Genome* **43**: 1011–1020.

LaChance, L. E. (1979). Genetic strategies affecting the success and economy of the sterile insect release method, pp. 8–18. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Lawson, E. T., T. A. Mousseau, R. Klaper, M. D. Hunter, and J. H. Werren (2001). Rickettsia associated with male-killing in a buprestid beetle. *Heredity* **87**: 497–505.

Lauge, G. (1985). Sex determination: genetic and epigenetic factors, pp. 298–317. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology*. Vol. (1). *Embryogenesis and Reproduction.* G. A. Kerkut, and L. I. Gilbert, Eds. Pergamon Press, Oxford.

Laurie, C. C. (1997). The weaker sex is heterogametic: 75 years of Haldane's Rule. *Genetics* **147**: 937–951.

Leslie, J. F. (1984). A "sex-ratio" condition in *Oncopeltus fasciatus*. *Heredity* **75**: 260–264.

Leventhal, E. (1968). The sex ratio condition in *Drosophila bifasciata*; its experimental transmission to several species of *Drosophila*. *J. Invert. Pathol.* **11**: 170–183.

Lewis, W. M. (1987). The cost of sex, pp. 33–57. In: *The Evolution of Sex and Its Consequences*. S. C. Stearns, Ed., Birkhauser Verlag, Basel, Switzerland.

Li, H., and B. S. Baker (1998). *Hermaphrodite* and *doublesex* function both dependently and independently to control various aspects of sexual differentiation in *Drosophila*. *Development* **125**: 2641–2651.

Lucchesi, J. C. (1999). On the origin of sex chromosomes. *BioEssays* **21**: 188–190.

Lyttle, T. W. (1993). Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends Genet.* **9**: 205–210.

Majerus, M. E. N., and G. D. D. Hurst (1997). Ladybirds as a model system for the study of male-killing symbionts. *Entomophaga* **42**: 13–20.

Majerus, T. M. O., J. H. G. von der Schulenburg, M. E. N. Majerus and G. D. D. Hurst (1999). Molecular identification of a male-killing agent in the ladybird *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae). *Insect Mol. Biol.* **8**: 551–555.

Makela, M. E., and M. D. Huettel (1979). Model for genetic control of *Heliothis virescens*. *Theor. Appl. Genet.* **54**: 225–233.

Marin, I., and B. S. Baker (1998). The evolutionary dynamics of sex determination. *Science* **281**: 1990–1994.

Marin, I., M. L. Siegal, and B. S. Baker (2000). The evolution of dosage-compensation mechanisms. *BioEssays* **22**: 1106–1114.

Maynard Smith, J. (1978). *The Evolution of Sex.* Cambridge University Press, Cambridge, UK.

McAllister, B. E., and G. A. T. McVean (2000). Neutral evolution of the sex-determining gene *transformer* in *Drosophila*. *Genetics* **154**: 1711–1720.

Meadows, M. E. (1985). Eradication program in the southeastern United States, pp. 8–11. In: *Symposium on Eradication of the Screwworm from the United States and Mexico.* O. H. Graham, Ed., Misc. Publ. (62). Entomol. Soc. Amer., College Park, MD.

Merrill, C., L. Bayraktaroglu, A. Kusano, and B. Ganetzky (1999). Truncated RanGAP encoded by the *Segregation Distorter* locus of *Drosophila*. *Science* **293**: 1742–1745.

Michod, R. E., and B. R. Levin, Eds. (1988). *The Evolution of Sex.* Sinauer Assoc. Sunderland, MA.

Miller, S. G., M. D. Huettel, M. B. Davis, E. H. Weber, and L. A. Weber (1986). Male sterility in *Heliothis virescens* X *H. subflexa* backcross hybrids. Evidence for abnormal mitochondrial transcripts in testes. *Mol. Gen. Genet.* **203**: 451–461.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* **1**: 2–9.

Naito, T., and H. Suzuki (1991). Sex determination in the sawfly, *Athalia rosae ruficornis* (Hymenoptera): occurrence of triploid males. *J. Hered.* **82**: 101–104.

Nothiger, R., and M. Steinmann-Zwicky (1985). A single principle for sex determination in insects, pp. 615–621. In: *Molecular Biology of Development*. Vol. (50). Cold Spring Harbor Symp. Quant. Biol. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Nur, U., J. H. Werren, D. G. Eickbush, W. D. Burke, and T. H. Eickbush (1988). A "selfish" B chromosome that enhances its transmission by eliminating the paternal genome. *Science* **240**: 512–514.

O'Neill, S. L., A. A. Hoffmann, and J. H. Werren, Eds. (1997). *Influential Passengers. Inherited Microorganisms and Arthropod Reproduction*. Oxford Univ. Press, Oxford, UK.

Palopoli, M., P. Doshi, and C. I. Wu (1994). Characterization of two *Segregation distorter* revertants: evidence that the tandem duplication is necessary for *Sd* activity in *Drosophila melanogaster*. *Genetics* **136**: 209–215.

Partridge, L., and L. D. Hurst (1998). Sex, and conflict. *Science* **281**: 2003–2008.

Pauli, D., and A. P. Mahowald (1990). Germ-line sex determination in *Drosophila melanogaster*. *Trends Genet.* **6**: 259–264.

Pineda-Vargas, N. (1985). Screwworm eradication in Mexico: activities of the Mexico–American screwworm eradication commission, 1977–84, pp. 22–27. In: *Symposium on Eradication of the Screwworm from the United States and Mexico.* O. H. Graham, Ed., Misc. Publ. (62). Entomol. Soc. Amer., College Park, MD.

Presgraves, D. C., E. Severance, and G. S. Wilkinson (1997). Sex chromosome meiotic drive in stalk-eyed flies. *Genetics* **147**: 1169–1180.

Randerson, J. P., N. G. C. Smith, and L. D. Hurt (2000). The evolutionary dynamics of male-killers and their hosts. *Heredity* **84**: 151–160.

Rao, S. R. V., and M. Padmaja (1992). Mammalian-type dosage compensation mechanism in an insect—*Gryllotalpa fossor* (Scudder)—Orthoptera. *J. Biosci.* **17**: 253–273.

Reed, K. M., and J. H. Werren (1995). Induction of paternal genome loss by the paternal-sex-ratio chromosome and cytoplasmic incompatibility bacteria (*Wolbachia*): a comparative study of early embryonic events. *Mol. Reprod. Dev.* **40**: 408–418.

Retnakaran, A., and J. Percy (1985). Fertilization and special modes of reproduction, pp, 213–293. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology*. Vol. (1). *Embryogenesis and Reproduction*. G. A. Kerkut, and L. I. Gilbert, Eds. Pergamon Press, Oxford.

Rice, W. R. (1994). Degeneration of a nonrecombining chromosome. *Science* **263**: 230–232.

Rice, W. R., and A. K. Chippindale (2001). Sexual recombination and the power of natural selection. *Science* **294**: 555–559.

Rigaud, T. (1999). Further *Wolbachia* endosymbiont diversity: a tree hiding in the forest? Trends Ecol. *Evol.* **14**: 212–213.

Rigaud, T., and F. Rousset (1996). What generates the diversity of *Wolbachia*–arthropod interactions? *Biodiversity Conserv.* **5**: 999–1013.

Robinson, A. S., and G. Franz (2000). The application of transgenic insect technology in the sterile insect technique, pp. 307–318. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Rokas, A. (2000). *Wolbachia* as a speciation agent. *Trends Ecol. Evol.* **15**: 44–45.

Rousset, F., and M. Raymond (1991). Cytoplasmic incompatibility in insects: why sterilize females? *Trends Ecol. Evol.* **6**: 54–57.

Shirk, P. D., D. A. O'Brochta, P. E. Roberts, and A. M. Handler (1988). Sex-specific selection using chimeric genes, pp. 135–146. In: *Biotechnology for Crop Protection*. P. A. Hedin, J. J. Menn and R. M. Hollingsworth, Eds. Am. Chem. Soc. Symp. Series No. 379. Washington, DC.

Shoemaker, D. D., V. Katju, and J. Jaenike (1999). *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria. Evolution* **53**: 1157–1164.

Sievert, V., S. Kuhn, A. Paululat, and W. Traut (2000). Sequence conservation and expression of the Sex-lethal homologue in the fly *Megaselia scalaris*. *Genome* **43**: 382–390.

Slee, R., and M. Bownes (1990). Sex determination in *Drosophila melanogaster*. *Quart. Rev. Biol.* **65**: 175–204.

Smith, E. R., A. Pannuti, W. Gu., A. Steurnagel, R. G. Cook, C. D. Allis and J. C. Lucchesi (2000). The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification liked to dosage compensation. *Mol. Cell. Biol.* **20**: 312–318.

Snow, J. W., C. J. Whitten, A. Salinas, J. Ferrer, and W. H. Sudlow (1985). The screwworm (*Cochliomyia hominivorax*) (Diptera: Calliphoridae) in Central America and proposed plans for its eradication south to the Darien Gap in Panama. *J. Med. Entomol.* **22**: 353–360.

Somerson, N. L., L. Ehrman, J. P. Kocka, and F. J. Gottlieb (1984). Streptococcal L-forms isolated from *Drosophila paulistorum* semispecies cause sterility in male progeny. *Proc. Natl. Acad. Sci. USA* **81**: 282–285.

Stevens, L. (1993). Cytoplasmically inherited parasites and reproductive success in *Tribolium* flour beetles. *Anim. Behav.* **46**: 305–310.

Stevens, L., R. Giordano, and R. F. Fialho (2001). Male-killing, nematode infections, bacteriophage infection, and virulence of cytoplasmic bacterial in the genus *Wolbachia*. *Annu. Rev. Ecol. Syst.* **32**: 519–545.

Stouthamer, R. (1997). *Wolbachia*-induced parthenogenesis, pp. 102–124. In: *Influential Passengers. Inherited Microorganisms and Arthropod Reproduction.* S. L. O'Neill, A. A. Hoffmann and J. H. Werren, Eds. Oxford Univ. Press, Oxford, UK.

Stouthamer, R., and R. F. Luck (1993). Influence of microbe-associated parthenogenesis on the fecundity of *Trichogramma deion* and *T. pretiosum*. *Entomol. Exp. Appl.* **67**: 183–192.

Stouthamer, R., and J. H. Werren (1993). Microbes associated with parthenogenesis in wasps of the genus *Trichogramma*. *J. Invert. Pathol.* **61**: 6–9.

Stouthamer, R., R. F. Luck, and W. D. Hamilton (1990). Antibiotics cause parthenogenetic *Trichogramma* (Hymenoptera/Trichogrammatidae) to revert to sex. *Proc. Natl. Acad. Sci. USA* **87**: 2424–2427.

Stouthamer, R., R. F. Luck, and J. H. Werren (1992). Genetics of sex determination and the improvement of biological control using parasitoids. *Environ. Entomol.* **21**: 427–435.

Stouthamer, R., J. A. J. Breeuwer, and G. D. D. Hurst (1999). *Wolbachia pipientis*: Microbial manipulator of arthropod reproduction. *Annu. Rev. Microbiol.* **53**: 71–102.

Thomson, M. S., and R. W. Beeman (1999). Assisted suicide of a selfish gene. *Heredity* **90**: 191–194.

Thornhill, R., and J. Alcock (1983). *The Evolution of Insect Mating Systems.* Harvard Univ. Press, Cambridge, MA.

Traut, W. (1994a). Sex determination in the fly *Megaselia scalaris*, a model system for primary steps of sex chromosome evolution. *Heredity* **82**: 163–169.

Traut, W. (1994b). Sex determination in the fly *Megaselia scalaris*, a model system for primary steps of sex-chromosome evolution. *Genetics* **126**: 1097–1104.

Traut, W., and U. Willhoeft (1990). A jumping sex determining factor in the fly *Megaselia scalaris*. *Chromosoma* **99**: 407–412.

Turelli, M., and A. A. Hoffmann (1999). Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Mol. Biol.* **8**: 243–255.

Vala, F., J. A. J. Breeuwer, and M. W. Sabelis (2000). *Wolbachia*-induced "hybrid breakdown" in the two-spotted spider mite *Tetranychus urticae* Koch. *Proc. R. Soc. London B* **267**: 1931–1937.

Vavre, F., F. Flury, J. Varaldi, P. Fouillet, and M. Bouletreau (2000). Evidence for female mortality in *Wolbachia*-mediated cytoplasmic incompatibility in haplodiploid insects: epidemiologic and evolutionary consequences. *Evolution* **54**: 191–200.

Vincent, S., L. A. Perkins, and N. Perrimon (2001). Doublesex surprises. *Cell* **106**: 399–402.

von der Schulenburg, J. H. G., M. Habig, J. J. Sloggett, K. M. Webberley, D. Bertrand, G. D. D. Hurst, and M. E. N. Majerus (2001). Incidence of male-killing *Rickettsia* spp. (α-Proteobacteria) in the ten-spot ladybird beetle *Adalia decempunctata* L. (Coleoptera: Coccinellidae). *Appl. Env. Microbiol.* **67**: 270–277.

Weeks, A. R., F. Marec, and J. A. J. Breeuwer (2001). A mite species that consists entirely of haploid females. *Science* **292**: 2479–2482.

Werren, J. H. (1997). Biology of *Wolbachia*. *Annu. Rev. Entomol.* **42**: 587–609.

Werren, J. H., and L. W. Beukeboom (1998). Sex determination, sex ratios, and genetic conflict. *Annu. Rev. Ecol. Syst.* **29**: 233–261.

Werren, J. H., U. Nur, and D. Eickbush (1987). An extrachromosomal factor causing loss of paternal chromosomes. *Nature* **327**: 75–76.

Werren, J. H., D. Windsor, and L. Guo (1995). Distribution of *Wolbachia* among neotropical arthropods. *Proc. R. Soc. Lond. B* **262**: 197–204.

Werren, J. H., G. D. D. Hurst, W. Zhang, J. A. J. Breeuwer, R. Stouthamer and M. E. N. Majerus (1994). Rickettsial related male-killing in the ladybird beetle (*Adalia bipunctata*). *J. Bacteriol.* **176**: 388–394.

West, S. A., E. A. Herre, B. C. Sheldon (2000). The benefits of allocating sex. *Science* **290**: 288–290.

White, M. J. D. (1973). *Animal Cytology and Evolution*. Cambridge Univ. Press, Cambridge, UK.

Whiting, P. W. (1943). Multiple alleles in complementary sex determination of *Habrobracon*. *Genetics* **28**: 365–382.

Wilkinson, G. S., and P. R. Reillo (1994). Female choice response to artificial selection on an exaggerated male trait in a stalk-eyed fly. *Proc. R. Soc. Lond. B* **255**: 1–6.

Wilkinson, G. S., D. C. Presgraves, and L. Grymes (1998). Male eye span in stalk-eyed flies indicates genetic quality by meiotic drive suppression. *Nature* **391**: 276–279.

Williamson, D. L., B. Sakaguchi, K. J. Hackett, R. F. Whitcomb, J. G. Tully, P. Carle, J. M. Bove, J. R. Adams, M. Konai and R. B. Henegar (1999). *Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. *Intern. J. Syst. Bacteriol.* **49**: 611–618.

Wood, R. J., and M. E. Newton (1991). Sex-ratio distortion caused by meiotic drive in mosquitoes. *Am. Naturalist* **137**: 379–391.

Wrensch, D. L., and M. A. Ebbert, Eds. (1993). *Evolution and Diversity of Sex Ratio in Insects and Mites.* Chapman and Hall, New York.

Wyss, J. H. (2000). Screw-worm eradication in the Americas—Overview, pp. 79–86. In: *Area-Wide Control of Fruit Flies and Other Insect Pests.* K. H. Tan, Ed. Penerbit Universiti Sains Malaysia, Pulau Pinang.

Xie, T., and A. C. Spradling (2000). A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* **290**: 2328–2330.

Zchori-Fein, E., R. T. Roush, and M. S. Hunter (1992). Male production induced by antibiotic treatment in *Encarsia formosa* (Hymenoptera: Aphelinidae), an asexual species. *Experientia* **48**: 102–105.

Zchori-Fein, E., Y. Gottlieb, S. E. Kelly, J. K. Brown, M. N. Wilson, T. L. Karr, and M. S. Hunter (2001). A newly discovered bacterium associated with parthenogenesis and a change in host selection behavior in parasitoid wasps. *Proc. Natl. Acad. Sci. USA* **98**: 12555–12560.

Some Relevant Web Sites

Sex determination in *Drosophila*: The Interactive Fly: http://sdb.bio.purdue.edu/fly/

The European *Wolbachia* Project: Towards Novel Biotechnological Approaches for Control of Arthropod Pests and Modification of Beneficial Arthropod Species by Endosymbiotic Bacteria:
http://wit.integratedgenomics.com/GOLD/Wolbachia.html

http://www.ncbi.nim.nih.gov

The screwworm *Cochliomyia hominovorax*:
www.nal.usda.gov/speccoll/collect/screwworm

11

Molecular Genetics of Insect Behavior

11.1. Overview

The study of insect behavior involves the analysis of any and all activities performed by an insect in relation to its surrounding environment. Behavior genetics is the study of the underlying hereditary basis of the behavior. For many years, Mendelian genetic analyses were conducted on a few traits determined by one or a few genes, or quantitative genetic methods were used for traits determined by “many” genes. The *Drosophila* Genome Project

and the use of molecular genetic methods are revolutionizing the analysis of the genetic basis of insect behavior.

Circadian behaviors, mating behavior, and learning in *Drosophila* have been dissected with the tools of molecular genetics, and inter- and intraspecific comparisons can be made of the DNA sequences associated with these behaviors. The circadian clock of *Drosophila* involves several genes, including *period*$^+$ (*per*$^+$). Mutants of *period* influence activity patterns and other circadian rhythms, as well as altering song cycles in courting males. The *per*$^+$ locus has been cloned and sequenced in *D. melanogaster* and *D. simulans*. After the *per*$^+$ gene of *D. simulans* was inserted by *P*-element-mediated transformation into a strain of *D. melanogaster* that is arrhythmic, transgenic *D. melanogaster* males produced song cycles like those of *D. simulans*. The differences in song rhythm maps to a small segment of the *per*$^+$ locus that may vary by as few as four amino acids. In addition, *timeless*$^+$, *doubletime*$^+$, *cycle*$^+$, *cryptochrome*$^+$, and *Clock*$^+$ are involved in the circadian clock. The clock involves transcription of the *per*$^+$ and *tim*$^+$ genes, followed by production of the PER and TIM proteins and subsequent negative feedback on self-transcription. Degradation of proteins then releases the negative feedback, allowing a new round of transcription, resulting in oscillations of RNA and protein.

Drosophila learning mutants, such as *dunce* and *couch potato*, are providing insights into the fundamental processes involved in short-term, intermediate, long-term, and anesthetic-resistant learning in insects and other organisms.

Analyses of behaviors that are determined by many genes are being revolutionized by the use of molecular genetic methods and the *Drosophila* Genome Project. It now is possible to map the number and location of genes affecting complex traits by correlating their inheritance with a variety of DNA markers and by conducting gene chip analyses.

11.2. Introduction

Insect behavior covers a very wide range of activities, including locomotion, grooming, feeding, communication, reproduction, dispersal, flight, learning, migration, host or prey selection, diapause, and various responses to environmental hazards such as temperature, humidity, parasites, and toxins (Dingle 1978, Beck 1980, Dingle and Hegmann 1982, Alcock 1984, Tauber et al. 1986, Gatehouse 1989, 1997, Sokolowski 2001). Understanding the behavior of pest and beneficial insects could allow one to improve pest management programs (Foster and Harris 1997, Renou and Guerrero 2000).

One definition of **behavior** is any action that an individual carries out in response to a stimulus or to its environment, especially an action that can be observed and described. However, insects also behave spontaneously, in the absence of any obvious stimulus. Thus, behavior includes studies to understand how an insect takes in information from its environment, processes that information, and acts. Processing information in the central nervous system may involve integrating information over time, including stimuli such as hormones coming from within the insect. Thus, the connection between **stimulus** and **response** can be delayed and indirect.

The genetic analysis of behavior rightfully has been perceived to be more complex than the analysis of morphological or anatomical traits (Baker et al. 2001, Sokolowski 2001).

One of the complications in genetic analyses of behavior is the difficulty of defining the behavior in a clear manner. Distinguishing between behavior and physiology is particularly difficult. The same behavior can be examined from at least four different viewpoints: 1) the immediate cause (or control); 2) its development during the individual's lifespan; 3) the function of the behavior; and 4) how the behavior evolved (Wyatt 1997).

Behavior genetics began to grow as a field of study in the 1960s, but was limited to demonstrating that a behavioral trait was heritable, determining whether its **mode of inheritance** was dominant or recessive, sex-linked or autosomal, and resolving whether the variation was due to single or multiple genes (Ehrman and Parsons 1973).

Genetic analyses of insect behavior require careful control of environmental conditions, because even subtle differences in test conditions can influence the results of assays. Objective measures of insect behavior often are difficult, and considerable efforts have been devoted to devising specific and appropriate assays. The possible influence of learning always must be considered, and, to complicate matters further, learning rates no doubt vary among different populations of the same species so that both heredity and environment must be considered. Genetic analyses of insect behavior involve, in many cases, analyses of the physiological or morphological changes that are associated with the change in behavior. Sometimes behavior is changed in an insect because a morphological trait has been altered through mutation.

The genetic basis of insect behavior has been analyzed most extensively using *Drosophila melanogaster* and a few other species, such as honey bees, grasshoppers, *Nasonia* parasitoids, and crickets (Benzer 1973, Matthews and Matthews 1978, Ehrman and Parsons 1981, Hall 1984, Kalmring and Elsner 1985, Huettel 1986, Huber et al. 1989, Menzel 1999, Beukeboom and van den Assem 2001).

Molecular genetic techniques provide powerful methods to analyze insect behavior such as olfaction, learning, circadian rhythms, and mating behavior. Having the complete sequence of the genome of *D. melanogaster* simplifies the isolation of specific genes that are involved in the behavior. *P*-element mediated transformation makes it possible to insert genes from one species of *Drosophila* into the genome of another, and their effect(s) on behavior can be determined. Transgenic *D. melanogaster* carrying markers such as green fluorescent protein (GFP) allow scientists to determine when and where specific genes are active.

Molecular genetic analyses of learning and memory in *Drosophila* may provide a means to study one of the most challenging frontiers in neurobiology (Waddell and Quinn 2001). Molecular genetic methods may allow us to localize and identify some of the individual genes among the "many" involved in determining the interesting and complex behaviors exhibited by insects (Doerge 2002).

Analyses of insect behavior employ techniques from several disciplines including anatomy, biochemistry, ecology, **ethology** (study of animal behavior in the natural environment), genetics, psychology, physiology, and statistics (Matthews and Matthews 1978, Hay 1985, Bell 1990, Holman et al. 1990, Via 1990, Barton Browne 1993, Heisenberg 1997, Doerge 2002). These disciplines are required because an insect perceives the environment through its sensory systems. The external sensory stimuli are transduced into electrical information, which is then processed and decoded, leading to a behavioral response. Behavior can be divided into several sequential steps: **stimulus recognition**, **signal transduction**, **integration**, and **response** or motor output.

11.3. The Insect Nervous System

The insect brain contains around 10^5 to 10^6 neurons. It consists of three main divisions: the **protocerebrum**, **deutocerebrum**, and **tritocerebrum**. In each of these divisions, different **neuropil** regions are located; a neuropil is a dense network of interwoven axons and dendrites of neurons and neuroglial cells in the central nervous system and parts of the peripheral nervous system.

In the protocerebrum, centers are present that are associated with vision and other sensory receptors (the mushroom bodies and central complex). The superior protocerebrum, with the pars intercerebralis, contains sets of neurosecretory cells that supply neurohemal organs in the corpora cardiaca and corpora allata, which are located in the head or prothorax in insects. The optic lobes flanking the protocerebrum contain the most well-organized neuropils in the brain.

Mushroom bodies in the brain are associated with olfactory pathways, including olfactory learning (Figure 11.1). Among the insects, mushroom bodies differ greatly in size and shape, with the number of cells ranging from 2500 in *Drosophila* to 50,000 in the cricket *Acheta*, 170,000 in the honey bee, and 200,000 in the cockroach *Periplaneta* (Heisenberg 1998, Strausfeld et al. 1998). The antennal centers are found in the deutocerebrum; in the tritocerebrum, neurosecretory neurons and neurons associated with the control of feeding and foregut activity are found (Homberg et al. 1989). The brain is connected to the subesophageal ganglion via connectives and to the thoracic and abdominal ganglia, or ventral nerve cord (Strausfeld 1976).



Figure 11.1. Photograph of the brain structures known as mushroom bodies in *D. melanogaster*. The dark areas show the mushroom bodies stained with an antibody to the *dunce*-encoded enzyme. *dunce* encodes an enzyme called cyclic AMP phosphodiesterase, which destroys cyclic AMP, which is important in learning and memory. (Photograph provided by R. L. Davis.)

Behavior and development are coordinated in the insect by both nerves and neuropeptides. Both neurosecretory cells and neurons use **neuropeptides** as messengers. Many different types of neuropeptides have been identified, including proctolin and adipokinetic hormone (Scharrer 1987, Masler et al. 1993, Raina and Menn 1993). Neuropeptides range in size from three amino acid residues (thyrotropin-releasing hormone) to more than 50 (insulin). They are generated from larger precursor proteins, ranging from 90 to 250 amino acids in length. A number of neuropeptide genes have been cloned, including bombyxin or prothoracicotropic hormone (PTTH), eclosion hormone (EH), FMRFamide-related peptides, diapause hormone, and pheromone biosynthesis-activating neuropeptide (PBAN) (Nassel 1993, Sato et al. 1993, Tillman et al. 1999).

Neuropeptides are released as cotransmitters and modulate fast transmission at neuromuscular junctions. A given neuropeptide may occur at several different sites, including central nervous system circuits, peripheral synapses, and at the peripheral targets (muscles and glands). Neuropeptides regulate behavior by coordinating the temporal and spatial activity of many neuronal circuits. Each of the circuits controlling behavior employs sets of sensory neurons, interneurons, and motor neurons. Thus, multiple neural networks share neural elements. Molecular genetic analysis is providing rapid progress in understanding neuropeptide receptors and second messenger pathways. Research on neuropeptides and their receptors indicates that they have roles during embryonic development and as cytokines in the immune systems of insects (Nassel 1993).

11.4. Traditional Genetic Analyses of Behavior

Sometimes mutations in a single gene or a few major genes will alter a behavior, and the mode of inheritance can be assessed by traditional methods. Traditional behavior-genetic analysis employs two main experimental approaches: **crossing** and **selection**. A third, limited to *D. melanogaster*, involves analysis of **fate maps** in genetic mosaics to locate the anatomical site of abnormalities that affect behavior (Hotta and Benzer 1972).

Although a specific behavior sometimes can be altered by the mutation of a single gene, an insect's behavior often is influenced by many genes (Plomin 1990). In such situations, analyses of behavior traditionally have required the use of **quantitative genetic** methods.

11.4.1. Crossing Experiments

A crossing experiment involves mating individuals that differ in a particular kind of behavior and then examining the behavior of their F_1 and backcross progeny. (A backcross is a cross of F_1 individuals to a parental line, usually the homozygous recessive one.) Ideally, the environment is controlled so that all individuals experience the same conditions. It is easiest to interpret the results of the experiment *if* the individuals that are crossed differ *only* with regard to a single behavioral attribute.

The phenotype of the F_1 and backcross progeny indicates whether the behavior is determined by a single gene or multiple genes, and whether there is dominance, sex linkage, or maternal influences. If the trait is determined by many genes, it is difficult to determine the number of loci, their relationship to each other, or their location on specific chromosomes because most insect species lack sufficient genetic markers. New molecular and statistical methods using quantitative trait loci may provide greater power to study and locate multiple and interacting loci in the future (Doerge 2002).

Table 11.1. Crossing Experiment to Explain Differences in Nest-Cleaning Behavior among Inbred Lines, F_1, and Backcross Progeny of the Haplo-diploid Honey bee *Apis mellifera* Supports the Two-Locus, Two-Allele Model

Parental lines	Hygienic (diploid) *uu*, *rr* queen		X ↓ ↓		Unhygienic u^+, r^+ (haploid) male	
F_1 progeny			All diploid sterile workers	u^+u, r^+r	(Unhygienic workers)	
			Backcross progeny (workers)			
Cross of u^+u, r^+r queen X hygienic *u*, *r* male	→	1: *uu*, *rr* Hygienic	1: *uu*, r^+r Uncaps, doesn't remove	1: u^+u, *rr* Removes, doesn't uncap	1: u^+u, r^+r Unhygienic	Ratio

Data from Rothenbuhler (1964).

One aspect of honey-bee behavior provides an example of a trait that is determined by a few genes. Other behaviors, including house-entering behavior in the mosquito *Aedes aegypti* and foraging in *D. melanogaster*, also appear to be determined by one or a few genes.

11.4.1.1. Susceptibility to American Foulbrood in *Apis mellifera*

Susceptibility to foulbrood disease caused by *Bacillus larvae* originally was analyzed by crossing two inbred honey-bee strains with differing levels of resistance (Table 11.1). The differences in resistance were attributed to differences in "hygienic behavior" in worker (sterile female) bees (Rothenbuhler 1964). Resistant workers (= hygienic) consistently removed dead larvae and pupae from the brood nest at a high rate, thus slowing the spread of the bacteria through the colony by reducing contamination. Crosses between hygienic queens and susceptible nonhygienic haploid males yield F_1 worker progeny that are nonhygienic, indicating that the genes conferring resistance are recessive.

Progeny produced by backcrosses to the homozygous recessive hygienic strain yielded approximately 25% hygienic workers, which is consistent with the hypothesis that hygienic behavior is determined by two recessive loci (Table 11.1). Under this two-locus model, hygienic queens are homozygous for two genes, *uu* and *rr*. The hygienic workers (*uu, rr*) both uncap the cells (*uu*) containing dead brood *and* remove them (*rr*).

Analysis indicated that *uu*, r^+r individuals will uncap the cells but not remove dead brood (Table 11.1). The u^+u, *rr* individuals do not uncap brood, but will remove them if the cells are uncapped for them. Individuals that are u^+u, r^+r are unhygienic and will neither uncap nor remove brood. Hygienic behavior appears to be a general response to remove pathogens and parasites from the nest (Spivak and Gilliam 1993).

Rothenbuhler's research on hygienic behavior became a classic in textbooks on behavioral genetics because it was one of the first and best examples that demonstrated that behavior was inherited (Rothenbuhler 1964). More recently, Moritz (1988) proposed a three-locus model for hygienic behavior in bees. Research on hygienic behavior continues because such behavior is important in managing bees. It is clear that the expression of hygienic behavior depends on colony strength and composition of worker types within the

colony (Spivak and Gilliam 1993, Arathi et al. 2000). Electro antennogram analyses of the olfactory and behavioral responses of hygienic and nonhygienic bees to diseased brood indicates that hygienic bees have a higher sensitivity to low concentrations of the odor of diseased bee pupae (Masterman et al. 2001). Such differences are due to a lower stimulus threshold and are not a direct result of age or experience of the bee. Thus, nonhygienic bees may be unable to detect diseased brood.

Understanding hygienic behavior in *A. mellifera* has resulted in practical recommendations to bee keepers for selecting colonies resistant to chalkbrood (a fungal disease) and the pest bee mite *Varroa*. So far, no negative effects have been found associated with hygienic colonies, and such colonies produce as much honey as nonhygienic ones (Moritz 1994, Spivak and Gilliam 1998).

11.4.1.2. House-Entering Behavior in *Aedes aegypti*

House-entering behavior by the mosquito *A. aegypti* from East Africa has been analyzed by crossing different populations with different behaviors (Trpis and Hausermann 1978). One population of *A. aegypti* commonly enters houses (domesticated or D), while others rarely do so (either peridomestic, P, or feral, F). House-entering behavior is important in determining whether a population transmits yellow fever to humans.

Three populations of *A. aegypti* collected either inside houses (D), near a village (P), or from tree holes in a forest (F) were bred in insectaries and crossed to produce hybrid (DP, PD, DF, FD, PF, FP) populations (Trpis and Hausermann 1978). The original and hybrid populations were then marked with different colored fluorescent powders and released near houses. Marked mosquitoes were captured inside houses and in the village area. Of the mosquitoes entering houses, 45% were from the domestic (D) population, 13.9% were from hybrids between the domestic and peridomestic population (DP and PD), 9.8% were from the peridomestic population (P), and 5.7% were hybrids between the domestic and feral populations (DF and FD). Only 1.5 and 0.6% of the PF and FP hybrids were collected in the house, and the feral population entered the house with a frequency of only 0.6%. The recapture rates in the village area were in the reverse order. The data indicate the behavior is determined by a few genes with additive effects.

"Domesticity" in *A. aegypti* is a complex phenomenon that includes a variety of behaviors, including a preference for ovipositing in man-made containers, the ability of larvae to develop in drinking water stored in clay pots with a low nutritional content, and preferences for feeding on man (rather than birds) inside houses, as well as resting and mating indoors. No doubt *A. aegypti* speciated long before man began to build houses, but *A. aegypti* has adapted rapidly to human habitats, and the domestic form of *A. aegypti* is the only one known that is entirely dependent on man (Trpis and Hausermann 1978).

11.4.1.3. Foraging in *Drosophila*

Drosophila melanogaster larvae feed on yeast growing on decaying fruit. Naturally occurring populations contain individuals that vary in the distance the larvae travel while foraging for food, a difference attributed to a single *foraging* gene (Osborne et al. 1997, Sokolowski et al. 1997). Natural populations comprise approximately 70% "rovers" (who forage long distances) and 30% "sitters" (short-distance foragers), with the rover phenotype dominant to sitter, indicating a single-gene mode of inheritance (deBelle and Sokolowski 1987, Sokolowski 2001).

Sitter larvae grow at a normal rate and are of normal size. Both sitters and rovers are maintained in the field by natural selection; density-dependent selection can shift allele frequencies so that rovers are selected for in crowded larval environments and sitters in less crowded ones. The *foraging* gene codes for a cyclic guanosine monophosphate (cGMP)-dependent protein kinase, and rovers have higher kinase activity than sitters. Thus, subtle differences in this kinase lead to naturally occurring variation in behavior (Shaver et al. 1998). Another gene, *Chaser*, also affects larval foraging by increasing foraging path length (Pereira et al. 1995).

11.4.1.4. Other Behaviors Influenced by One or a Few Genes

Crossing experiments have shown that a specific behavior is influenced by one or a few genes in the flour moth *Ephestia kuhniella* (silk mat spinning by larvae prior to pupation), the mosquito *Aedes atropalpus* (egg maturation without an exogenous source of protein such as blood), and the parasitoid wasp *Habrobracon juglandis* (flightlessness) (Ehrman and Parsons 1981). In the silkworm *Bombyx mori*, females with the *piled egg* gene deposit eggs in a peculiar manner. *B. mori* larvae with the *Non-preference* gene are unable to discriminate mulberry leaves from others (Tazima et al. 1975), and Huettel and Bush (1972) found that when two monophagous tephritid flies (*Procecidochares*) were crossed, the host preference behavior segregated in a manner consistent with control by a single locus.

A variety of mutants determined by single major genes have been identified in *D. melanogaster* that affect behavior (Grossfield 1975, Hall 1985, Pavlidis et al. 1994), including a group of sex-linked, incompletely dominant mutants (*Shaker*, *Hyperkinetic*, and *ether-a-go-go*) that are expressed when the flies are anesthetized with ether. The *Hyperkinetic* mutants exhibit a vigorous steady leg shaking, while mutations at the *Shaker* locus cause vigorous and erratic shaking and a strong scissoring of wings and twitching of the abdomen. The *ether a-go-go* mutant flies are less vigorous in their shaking. The *easily shocked* gene of *D. melanogaster* is one of the "bang-sensitive" paralytic genes. Flies with this mutated gene exhibit a transient paralysis following a brief mechanical shock (Pavlidis et al. 1994). Sex-linked temperature-sensitive recessive mutant $para^{ts}$ causes *D. melanogaster* to become immobile above 29°C. Mutants of the *couch potato* locus cause flies to be hypoactive and exhibit abnormal geotaxis (response to gravity), phototaxis (response to light), and flight behavior. This gene is unusually complex, spanning more than 100 kb and encoding three different messages (Bellen et al. 1992).

Many "single gene" mutants that affect the morphology of *D. melanogaster* also affect behavior. Some mutant flies exhibit abnormal behavior because they are unable to perform the reaction to a stimulus due to altered effector structures. Other mutants exhibit altered behavior because perception of cues is impaired. For example, flies with *white* eyes may exhibit abnormal courtship behaviors (Grossfield 1975).

Pheromone communication in the European corn borer *Ostrinia nubialis* is genetically determined (Klun and Maini 1979, Klun and Huettel 1988, Lofstedt et al. 1989, Lofstedt 1990). Females of the E- and Z-strains of *O. nubialis* produce different **enantiomeric** ratios of sex pheromone. Hybrids between these two strains produce an intermediate pheromone blend. Analysis of the F_2 and backcross progeny indicates pheromone type is controlled by two alleles at a single autosomal locus, although one or more modifier genes controls the precise ratio of the isomers in heterozygous females.

Males of the two *O. nubialis* strains are attracted to the appropriate pheromone blends in the field, and hybrid males respond preferentially to the pheromone produced by

heterozygous females rather than to the pheromones produced by the two parental female types. The response of males to the pheromone is determined by a single sex-linked gene with two alleles. The olfactory sensillae of the two types of males are different, which is controlled by an autosomal locus with two alleles. Hybrid males give intermediate results when tested for their electrophysiological responses, with E- and Z-cells yielding approximately equal spike amplitudes. The genes determining variation in pheromone production and organization of male olfactory sensillae are not closely linked and are probably on different chromosomes (Lofstedt 1990).

11.4.2. Selection Experiments

Selection experiments provide another traditional method to determine the degree to which a given behavior is determined genetically. In a selection experiment, individuals with a specific behavioral attribute are allowed to reproduce, and this process is repeated over succeeding generations. Eventually, the behavior of the selected population is altered *if* genetic variation for the attribute is present in the initial colony *and* the selection procedures have been appropriate. The response of the population to selection can be analyzed to estimate the heritability of the trait.

11.4.2.1. Migratory Behavior in *Oncopeltus fasciatus*

For example, migratory behavior of the large milkweed bug, *Oncopeltus fasciatus*, was demonstrated to be under genetic control (Palmer and Dingle 1989). Strains of *O. fasciatus* were selected for wing length and propensity to fly. Bidirectional selection (selection for increased and decreased wing length) was performed for 13 generations, and the flight behavior of individuals monitored. Individuals also were selected for flight time, and those whose flight times totaled 30 min were considered “fliers,” while those with a short flight time were labeled “nonfliers.”

Response to selection on wing length was rapid, and flight tests of the long- and short-winged insects indicated there was a positive correlation between wing length and flight duration. Selection after two generations for flight or nonflight likewise resulted in divergent responses, indicating a large genetic component to flight behavior.

11.4.2.2. Analysis of Selection Experiments

To estimate the degree of genetic influence on a specific behavior, two measures are used: the selection differential and the estimate of heritability. The **response to selection (R)** is the difference in mean phenotypic value between the offspring of the selected parents and the mean phenotypic value of the entire parental generation before selection (Falconer 1989).

$$R = h^2 S$$

R is the improvement or response to selection, h^2 is the heritability of the characteristic under selection in the population, and S is the selection differential. The **selection differential (S)** is the average superiority of the selected parents expressed as a deviation from the population mean (Falconer 1989). The selection differential measures the difference between the average value of a quantitative character in the whole population and the average value of those selected to be parents of the next generation. It is measured in standard deviation units.

Heritability *in the broad sense* is the degree to which a trait is genetically determined. Because behavioral traits are influenced by both genes and environment, heritability is expressed as the ratio of the total genetic variance to the phenotypic variance (V_G/V_P). Heritability *in the narrow sense* is the degree to which a trait is transmitted from parents to offspring and is expressed as the ratio of the additive genetic variance to the total phenotypic variance (V_A/V_P) (Falconer 1989).

Heritability could be estimated to be zero if the specific population being selected had no variability for the behavioral attribute under study because it was inbred. Heritability could be estimated to be one if the trait was completely determined by genes, and the environment had little effect on the phenotype, although this would be an unusual outcome.

Heritability estimates provide no information about the actual mode of inheritance of a quantitative trait because they represent the cumulative effect of all loci affecting the trait. The number of loci involved generally can be determined only with elaborate and specially designed experiments.

A number of assumptions are made when estimating heritability: 1) all loci affecting the trait act *independently* of one another, and 2) the loci are *unlinked* (located on different chromosomes). Another assumption 3) is that environment affects all genotypes in a similar fashion. These three assumptions are not always justified. Thus, heritability estimates are difficult to interpret, although they are useful for predicting response to selection under specific environmental conditions.

Heritability estimates usually are made by regression-correlation analyses of close relatives (parent–offspring, full sibs, half sibs), experiments involving response to selection, or analysis of variance components. Traits with high heritabilities respond readily to selection with an appropriate selection method. The magnitude of the response to the selection, that is, the differences in mean values between parent and progeny generations, provides an estimate of heritability in the narrow sense (h_n^2). This estimate is valid only for the population being examined, under the test conditions employed, for the behavior observed, and for the method of measurement employed.

Heritability of most insect behaviors is relatively high, probably because many arthropod behaviors are highly stereotyped (Ehrman and Parsons 1981). For example, the heritability of locomotor activity of *D. melanogaster* has been estimated to be 0.51, and the heritability of mating speed of male *D. melanogaster* has been estimated to be 0.33. Heritability for honey production from honey bees ranged from 0.23 to 0.75, depending upon the experimental conditions and colonies tested (Rinderer and Collins 1986). Italian honey bees are less able to remove the parasitic mite *Varroa* than Africanized bees, and the heritability of this ability was estimated to be 0.71 (Moretto et al. 1993). The inheritance of honey-bee stinging behavior and body size was investigated by RAPD-PCR markers using crosses of European honeybee queens and drones from an African bee. A significant effect was found for the tendency to sting and five quantitative trait loci (Hunt et al. 1998). Heritability of the length of the pre-reproductive period in *Helicoverpa armigera*, which is when migratory flight occurs in this noctuid moth, ranged from 0.54 to 0.16 (Colvin and Gatehouse 1993). Heritability of host selection behavior by *Asobara tabida*, a parasitoid of *Drosophila subobscura*, ranged from 0.03 to 1.0 depending upon the test method employed (Mollema 1991).

11.4.3. Some Polygenically Determined Behaviors

Behavior is often a continuous variable, controlled by multiple genes with small additive effects (Plomin 1990, Heisenberg 1997). With such behaviors, the task of teasing apart

the respective roles of genes and environment requires statistical analysis (Doerge 2002). *Drosophila* behaviors determined by multiple genes include locomotor activity, chemotaxis, duration of copulation, geotaxis, host plant preference, mating speed, phototaxis, preening, and the level of sexual isolation within and between species. Multiple genes influence host plant adaptation and host preference in insects, and learning also may affect host preference (Papaj and Prokopy 1989, Via 1990). Host plant choice usually is a hierarchy of several components. For example, attraction to a site from a distance and oviposition site preference (egg laying at the site) are genetically distinct in *Drosophila tripunctata* (Jaenike 1986).

The genetic basis of host-plant specialization in the fruit flies *Drosophila sechellia* and *D. simulans* is determined by a minimum of three or four loci that affect egg production, survival, and host preference (R'Kha et al. 1991). *Drosophila sechellia* breeds in a single plant, *Morinda citrifolia*, which is toxic to other *Drosophila* species. Its sympatric relative, *D. simulans*, breeds on a variety of plants. The two species can be crossed, and the F_1 hybrid embryos produced by *D. simulans* females are susceptible to *Morinda* fruit because susceptibility is maternally inherited and fully dominant.

D. sechellia is stimulated by *Morinda* to produce eggs, but oviposition in *D. simulans* is inhibited by this plant. In hybrid progeny, the inhibition observed in *D. simulans* is dominant. F_1 hybrids and backcross progeny exhibit intermediate, approximately additive, behavior. These differences result in isolation of the two species in nature, although their ranges overlap geographically. Thus, their ecological niches are determined by tolerance to toxic products in the ripe *Morinda* fruit, with *D. sechellia* exhibiting a strong preference for *Morinda*, an ability to detect fragrant volatiles from *Morinda* over a long distance, and a stimulation of egg production by *Morinda*. By contrast, egg production in *D. simulans* is inhibited by *Morinda*.

Other specific behavioral attributes that are inherited in a complex manner include: *Musca domestica* (number of attempts to mate by males); *Phormia regina* (high and low ability to learn to extend the proboscis to a stimulus applied to the forelegs); hybrid crickets (call rhythm of males; female response to calling songs); *Anopheles albimanus* (ability to avoid pesticides); *Apis mellifera* (high and low collection of alfalfa pollen, and stinging behavior) (Ehrman and Parsons 1981, Hall 1985, Rinderer 1986). Gould (1986) found that the propensity for cannibalism by larvae of *Heliothis virescens* is polygenically determined. Most of these behaviors were analyzed by selection experiments.

11.5. Molecular Genetic Analyses of Insect Behavior

Molecular genetic analyses are providing significant advances in our knowledge of behavior (Plomin 1990, Sokolowski 2001, Doerge 2002). Molecular genetic methods are unlikely to replace traditional methods of behavior analysis, but the ability to identify, clone, and sequence specific genes makes it easier to understand several behaviors, including the periodicity of biological rhythms, mating behavior, locomotion, and learning. It is now possible to clone a gene from one *Drosophila* species, insert it into a *P*-element vector, and introduce the exogenous gene into mutant strains of *D. melanogaster* to confirm that the putative gene does, in fact, code for the behavior of interest. Cloned genes from *Drosophila* can, in some cases, be used as probes to identify genes from other arthropods, which then can be sequenced and compared. The availability of the complete genome of *D. melanogaster* will allow analyses of behavior that could not be conducted previously, as will be described in the discussion of olfaction in *D. melanogaster*.

11.5.1. The Photoperiodic Clock

The potential that molecular genetics offers is exemplified by the analyses conducted using the *period*$^+$ and other clock genes of *D. melanogaster* (Table 11.2). Most insects exhibit particular behaviors at a specific time of the day, which are due to the action of a **circadian clock** that allows the insect to measure time (Kyriacou 1990, Takahashi 1992). Such circadian rhythms have a number of characteristics:

1. The clocks that regulate such behavior usually are "free running" in constant environments and are not simple responses to changes in light or temperature.
2. Although the rhythms are free running, an initial environmental signal is required to start the clock. Among the cues that "set" the clock are alternating light and dark cycles, high and low temperature cycles, or short pulses of light.

Table 11.2. Some Genes Involved in the Circadian Clock of *Drosophila melanogaster*

Gene (abbreviation)	Mutant phenotype(s)	Function(s)
period$^+$ (*per*$^+$)	Short-period, long-period; and arrhythmic flies	Negative transcription element; *per*$^+$ mRNA levels rise late in the day
	Affects locomotion, eclosion, courtship rhythms	Activated by the heterodimer of the CLOCK and CYCLE proteins
		PER proteins feed back negatively on their own transcription
		PER contains a protein dimerization domain called PAS
timeless$^+$ (*tim*$^+$)	Short-period, long-period, and arrhythmic files	Negative element; *tim*$^+$ mRNA levels rise late in the day
		TIM protein destabilized by light
	Affects locomotion, eclosion, sleep	TIM proteins feed back negatively on their own transcription, interact with PER
Clock$^+$ (*Clk*$^+$)	Arrhythmic	CLK, in combination with CYC, activates transcription of *per*$^+$ and *tim*$^+$
	Affects locomotion, eclosion, rhythm	CLK negatively regulates itself
		Mutants blind for "lights-on" response
cycle$^+$ (*cyc*$^+$)	Arrhythmic	CYC, in combination with CLK, activates transcriptions of *per*$^+$ and *tim*$^+$
	Affects locomotion and eclosion	CYC negatively regulates itself Mutants respond poorly to light–dark cycles
cryptochrome$^+$ (*cry*$^+$)	Photoreceptor	Sequence homologous to a photolyase; binds TIM in a light-dependent manner
	Resets rhythms	Altered light response in mutants
doubletime$^+$ (*d*$^+$)	Lengthens cycle in constant darkness; affects locomotion, sleep	Casein kinase I involved in phosphorylating PER, rendering it unstable in absence of TIM

Adapted from Emery et al. (2000), Sancar (2000), Harmer et al. (2001), Merrow and Roenneberg (2001), Sokolowski (2001), Williams and Sehgal (2001).

3. The circadian rhythm is relatively insensitive to changes in temperature (temperature compensated).
4. The clock can be reset by altering the cues that entrain the clock.

Drosophila melanogaster born and reared in constant darkness exhibit circadian locomotor activity rhythms as adults. However, the rhythms of the individual flies in these populations are not synchronized with one another (Sehgal et al. 1992). Rhythms can be synchronized if dark-reared flies are exposed to light treatments as first-instar larvae (or as later instars). Light treatments occurring prior to hatching of the first-instar larvae fail to synchronize adult locomotor activity rhythms, indicating the clock functions continuously from the time larvae hatch until adulthood. The rhythm can be advanced, delayed, or unchanged, depending on the phase of the cycle at which the cue is given.

The circadian rhythm has an approximate periodicity of 24 h. Molecular genetic analyses of *Drosophila* clock mutants are providing a fundamental understanding of the mechanisms of the circadian clock. Rapid advances have been made in the past few years in understanding the molecular aspects of circadian clocks in a variety of organisms. Circadian rhythms are found in all organisms and probably evolved early; common genetic elements are present in *Drosophila*, *Neurospora*, mammals, and cyanobacteria (Lakin-Thomas 2000, Loudon et al. 2000, Allada et al. 2001, Harmer et al. 2001, Merrow and Roenneberg 2001, Williams and Sehgal 2001). In *Drosophila*, the genes *period*$^+$, *timeless*$^+$, *Clock*$^+$, *cycle*$^+$, *doubletime*$^+$, and *cryptochrome*$^+$ are now known to be involved in the circadian clock (Lakin-Thomas 2000, Emery et al. 2000, Table 11.2).

Numerous reviews have compared the molecular, genetic, and neurological components of biological rhythms, reflecting the excitement of the scientific community in understanding the molecular basis of this complex behavior (Hall 1995, 1998a; Iwasaki and Thomas 1997, Dunlap 1998, 1999; Young 1998, Ishida et al. 1999, Giebultowicz 2000, Lakin-Thomas 2000, Wager-Smith and Kay 2000, Allada et al. 2001, Harmer et al. 2001, Williams and Sehgal 2001). The numerous reviews are nearly overwhelming, and Hall (1998b) questioned how it is possible "to review an over-reviewed subject—one whose reviews have even been reviewed."

11.5.1.1. The *period*$^+$ Locus of *Drosophila*

The *Drosophila per*$^+$ locus is on the X chromosome, and mutations of it influence eclosion, locomotor activity, and the length of the interpulse interval of the courtship song (Table 11.2). Eclosion of wild-type flies (emergence of adults from the pupal case) typically occurs around dawn, when the presence of dew and high relative humidity increases their survival rate (Figure 11.2A). Locomotor activity then decreases during midday and is followed by increased activity again in the evening. Three classes of mutant alleles exist; they shorten (*per*^{S} mutants have 19-h eclosion rhythms), lengthen (*per*^{L} mutants have 29-h eclosion rhythms), or completely abolish circadian eclosion and locomotor activity rhythms (*per*^{0} mutants). Flies with the *per*^{0} mutation eclose arrhythmically, but periodicity in eclosion can be restored by *P*-element-mediated transformation of arrhythmic flies using the wild-type *per*$^+$ allele (Bargiello et al. 1984; Figures 11.2B, 11.2C).

The *per*$^+$ gene is approximately 7 kb long and encodes a 4.5-kb transcript with eight exons, the first of which is noncoding (Figure 11.3). One of the most striking features of the protein is a series of threonine-glycine (Thr-Gly) repeats in the middle of the gene (Costa et al. 1992, Guantieri et al. 1999). The region encoding the Thr-Gly repeat is

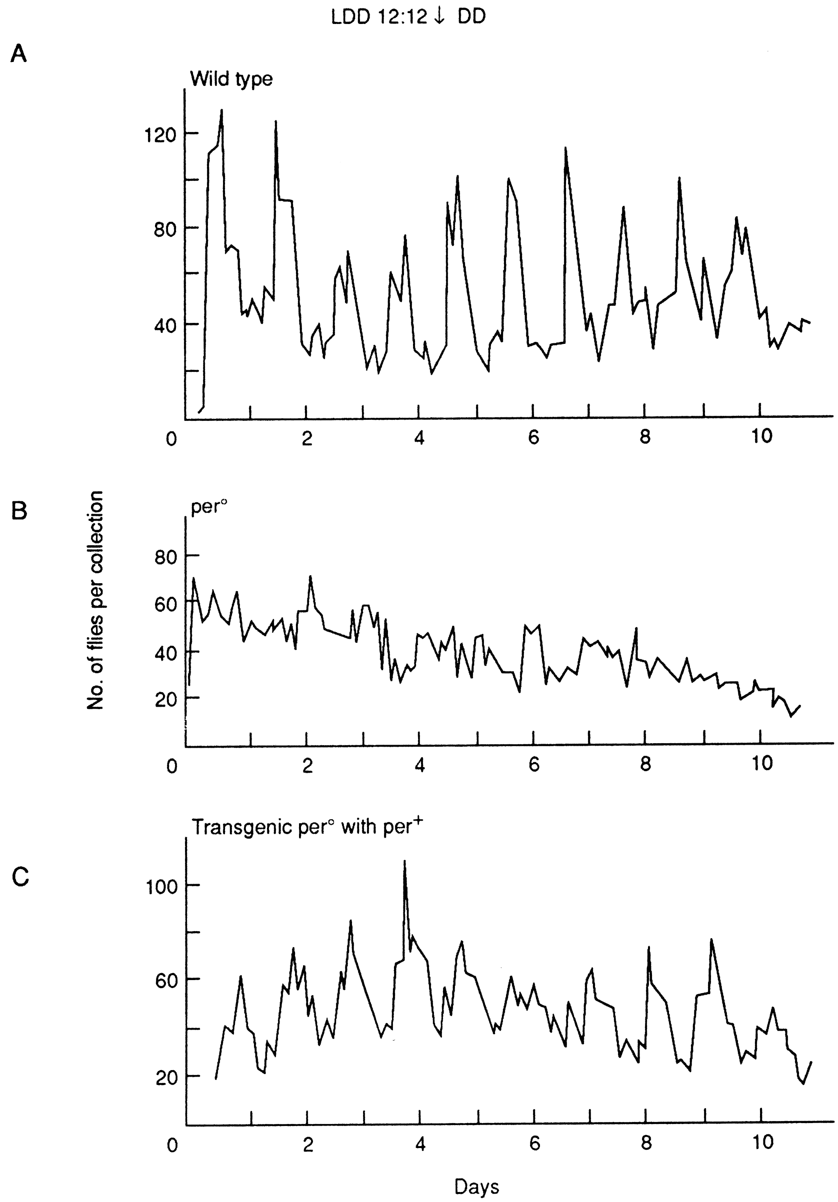


Figure 11.2. Profiles of eclosion (emergence of adults from pupal cases) for populations of A) *D. melanogaster* wild-type females , B) *per*^{0} males and females, and C) transgenic *per*^{0} individuals that have received a wild-type *per*$^+$ gene by P-element-mediated transformation. (Modified from Bargiello et al. 1984.)

polymorphic in length within and between *Drosophila* species and plays a role in the thermal stability of the circadian phenotype. For example, 17, 20, or 23 repeats are found in *D. melanogaster* populations, and a clinal pattern occurs along a north–south axis in Europe and North Africa, with the shorter sequences in southern Europe (Costa et al. 1992).

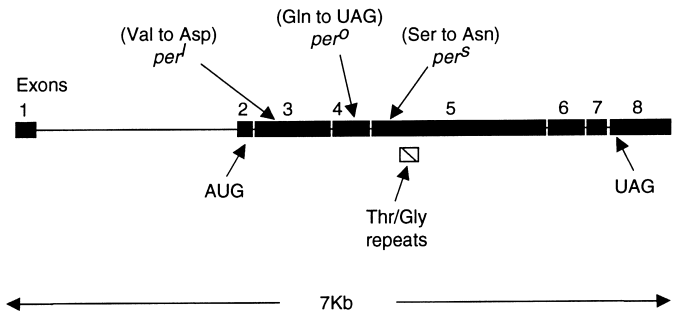


Figure 11.3. The exon/intron structure of the *D. melanogaster per* gene. The gene is approximately 7 kb long, with 7 exons. The locations of the per^L, per^0, and per^S mutations are indicated, as is the region which codes for the variable number (17, 20, or 23) of Thr-Gly repeats.

Costa et al. (1992) suggested that the length polymorphism cline is maintained by natural selection under different temperature conditions.

A large number of tissues express the per^+ product, including embryonic, pupal, and adult nervous systems, as well as the esophagus, gut, and ovaries. Liu et al. (1992) demonstrated that the per^+ gene product (PER) is predominantly found in cell nuclei in adult *Drosophila*, and Hardin et al. (1992) showed that per^+ mRNA levels undergo daily fluctuations, which constitutes a **feedback loop** in which PER affects the oscillations of its own mRNA. The fluctuations in per^+ mRNA are due to fluctuations in gene transcription because the per^+ mRNA has a relatively short half-life (Zerr et al. 1990), which is consistent with the hypothesis that PER acts as a transcription factor (Table 11.2).

The per^+ genes from *D. simulans*, *D. virilis*, *D. pseudoobscura*, and *D. yakuba* have been cloned and sequenced. Parts of the gene are conserved among them and parts are highly diverged, which suggests that conserved regions may encode basic functions common to all (clock-type functions), while species-specific differences such as love songs, locomotor activity, and eclosion profiles may be encoded within the variable regions (Kyriacou 1990).

$Clock^+$, $timeless^+$, $cycle^+$, and $doubletime^+$ are components of the circadian clock (Kyriacou 1993, Table 11.2). In addition, autosomal mutations induce flies to eclose early in a light–dark cycle; flies with mutations of $phase\text{-}angle^+$ emerge in the predawn part of the cycle instead of just after dawn, while flies with mutations of $gate^+$ fail to eclose during this narrow time window.

$cryptochrome^+$ (cry^+) is an important clock gene because it encodes a critical circadian photoreceptor in *Drosophila* (Egan et al. 1999, Emery et al. 2000). The gene product, CRY, belongs to a family of blue light-sensitive proteins which includes photolyases and plant blue light photoreceptors. Flies overexpressing CRY are hypersensitive to light. The CRY protein is probably the only dedicated circadian photoreceptor in *Drosophila* (Emery et al. 2000).

11.5.1.2. Song Cycle Behavior in Transgenic *Drosophila*

The courtship song is produced when males vibrate their wings. The song consists of two components: 1) courtship hums, and 2) a series of pulses with interpulse intervals, which can fluctuate between 15 and 85 milliseconds (Kyriacou and Hall 1989). The interpulse intervals have a period of 56 sec in *D. melanogaster* and 35 to 40 sec in *D. simulans*.

D. melanogaster males with the *per*^{S} mutation sing with 40-sec periods, *per*^{L} males sing with 76-sec periods, and *per*^{0} males are arrhythmic.

The genetic basis of species-specific song instructions was confirmed by the transfer of the *per*$^+$ gene cloned from *D. simulans* into *D. melanogaster* via *P*-element-mediated transformation (Wheeler et al. 1991). The *D. simulans per*$^+$ gene restored a rhythm in *D. melanogaster* lacking a rhythm, and transgenic *D. melanogaster* males produced song cycles characteristic of *D. simulans* males. Wheeler et al. (1991) concluded that substitutions in four or fewer amino acids in the *per*$^+$ locus are responsible for the species-specific courtship behavior.

11.5.1.3. Other Effects of *per*$^+$

The *per* alleles affect locomotion, cellular rhythms, and development time. Flies with *per*^{S} develop faster than wild-type flies, and *per*^{L} flies develop more slowly than the wild type (Kyriacou 1990, 1993).

It has long been thought that circadian oscillations provided the clock for photoperiodically induced **diapause** in insects (Takeda and Skopik 1997). Diapause is a genetically determined state of arrested development that is induced prior to the onset of detrimental conditions. Hibernal diapause, which allows insects to survive over winter, is often induced when insects develop during a period of cool temperatures under a short daylength, which means they must be able to measure light and dark cycles. However, *per*$^+$ appears to have no influence on the photoperiodic clock in *D. melanogaster* (Saunders 1990). Females of a wild-type strain of *D. melanogaster* (Canton-S) and strains with *per* mutations were able to discriminate between diapause-inducing short days and noninductive daylengths. *D. melanogaster* adult females exhibit an ovarian diapause when reared and held under short days and low temperature (12°C). Females exposed to long days at the same temperature reproduce. The critical daylength (the photoperiod at which 50% of the individuals enter diapause) for Canton-S females at 12°C is approximately 14 h of light per 24 h. Photoperiodic response curves for the *per*^{S}, *per*^{L}, and Canton-S strains were almost identical, although *per*^{0} flies showed shortened critical daylengths. However, *per*^{0} females are able to discriminate between a long day and a short day.

Many behaviors, including learning, involve temporally patterned events. The interval between presentation of the conditioned stimulus and reinforcement is important in associative learning. The conditioned stimulus must be presented before the unconditioned stimulus, and the unconditioned stimulus must follow the conditioned within a relatively brief interval. It was thought that the *per*$^+$ gene could be involved in learning, based on the observation that males with the *per*^{L} allele in one experiment did not exhibit normal experience-dependent courtship behavior. However, males with the wild-type or *per*^{S} and *per*^{0} alleles could be conditioned normally (Gailey et al. 1991).

11.5.2. Learning in *Drosophila*

It is difficult to produce a single definition of learning (Meller and Davis 1996, Tully 1996, Waddell and Quinn 2001). **Learning** can be defined as a change in behavior with experience, but this definition would not exclude responses such as growth and maturation, or other processes that are triggered by events such as mating or feeding. Another definition is a reversible change in behavior with experience, but this excludes phenomena in which the

modification caused by some experience is fixed and resistant to further change. Another definition is that learning is a more or less permanent change in behavior that occurs as a result of practice, but this definition is ambiguous (Papaj and Prokopy 1989).

Papaj and Prokopy (1989) suggest the following properties are characteristic of learning in insects: 1) The individual's behavior changes in a repeatable way as a consequence of experience. 2) Behavior changes gradually with continued experience, often following a "learning curve" to an asymptote. 3) The change in behavior accompanying experience declines in the absence of continued experience of the same type or as a consequence of a novel experience or trauma.

Insect populations vary in their ability to learn. Genetic variability within strains has been used to analyze learning in *Drosophila*, *Phormia* flies, and the honey bee (McGuire and Hirsch 1977, McGuire 1984, McGuire and Tully 1987, Tully 1996, Menzel 1999). *Drosophila melanogaster* can be sensitized and habituated, learn associations with positive or negative reinforcement, and be classically conditioned (Davis and Dauwalder 1991). *Drosophila melanogaster* can learn to run away from specific odors that they previously experienced with an electric shock, and hungry flies can learn to run toward odors previously associated with a sugar reward. Flies can learn visual, tactile, spatial, and proprioceptive cues (Waddell and Quinn 2001). Analyses of memory mutants in *Drosophila*, including *dunce*, *rutabaga*, *amnesiac*, *radish*, *zucchini*, *cabbage*, *tetanic*, *turnip*, *linotte* and *latheo*, indicate that memory consists of distinct phases: short-term, intermediate, long-term, and anesthesia-resistant memory (Table 11.3, Davis 1996, Sokolowski 2001).

Genetic analyses of learning in *D. melanogaster* began in the mid-1970s in Seymour Benzer's laboratory when *D. melanogaster* was trained to avoid an odor associated with a shock (Benzer 1973). The learned avoidance lasted only a few hours, but the odor avoidance test was used to screen mutagenized flies for strains that had normal olfaction and aversion to shock, but an abnormally low ability to associate odors with shocks. The mutant flies obtained were poor learners, but each had different phenotypes (Table 11.3). Flies with the mutant gene *amnesiac* had a nearly normal learning ability but forgot rapidly. Flies with mutated *dunce* genes had a shortened memory for several different conditioned behaviors (Davis and Dauwalder 1991) due to a defective gene for cAMP-specific phosphodiesterase, an enzyme that regulates levels of cyclic AMP (cAMP). The *dunce* flies have elevated cAMP levels (Zhong and Wu 1991). cAMP is part of a "second messenger" signaling pathway in nerve cells that help form associative memories. The *dunce* flies have impaired synaptic transmission because the excess of cAMP leads to hyperpolarization of the synaptic terminals, resulting in a chronically lowered availability of neurotransmitter (Delgado et al. 1992).

The *dunce*$^{+}$ gene is one of the largest and most complex identified in *Drosophila*, extending over 140 kb. It produces, by the use of multiple transcription start sites, alternative splicing of exons, and differential processing of 3$'$ sequences, at least eight to 10 RNAs ranging in size from 4.2 to 9.5 kb. One unusually large intron, 79 kb in length, contains at least two genes (*Sgs-4* and *Pig-1*) within it (Chen et al. 1987, Qiu et al. 1991). This "genes within an intron" arrangement is uncommon. One of the contained genes, *Sgs-4*$^{+}$, is expressed in larval salivary glands and provides the glue used by larvae to attach themselves to the surface for pupation. *Sgs-4*$^{+}$ is transcribed in the same direction as *dunce*$^{+}$. The second gene, *pre-intermolt*$^{+}$, also is expressed in larval salivary glands but is transcribed in the opposite direction. Genes homologous to *dunce*$^{+}$ have been identified in mice, rats, and humans, and the mammalian counterpart of *dunce*$^{+}$ functions in regulating mood (Tully 1991a).

Table 11.3. Some Single Genes Involved in Learning and Memory of *Drosophila melanogaster*

Gene	Mutant phenotype(s)	Function(s)
dunce^{+}	Short-term memory defective	cAMP-specific phosphodiesterase that degrades the adenylate cyclase produced by *rutabaga*$^{+}$ Affects locomotor rhythms, ethanol tolerance, learning
rutabaga^{+}	Short-term memory defective	Adenylate cyclase decreases expression of cAMP, affects courtship, learning, ethanol tolerance, grooming
amnesiac^{+}	Middle-term memory defective	Neuropeptide, stimulates cAMP synthesis Affects ethanol tolerance
radish^{+}	Anesthesia-resistant long-term memory	Affects only one type of long-term memory
cabbage^{+}	Long- and short-term memory (?)	
turnip^{+}	Long- and short-term memory	Involved in the protein kinase C pathway Affects olfactory discrimination and larval, visual, and reward learning
latheo^{+}	Acquisition of initial memory defective?	Involved in short-term memory; affects DNS replication and synaptic plasticity?
linotte^{+}	Retarded learning Mutants have structural brain defects (mushroom bodies and central complex)	Encodes a novel protein or is an allele of the *derailed* receptor tyrosine kinase
Volado^{+}	Short-term memory Expressed in mushroom	Cell surface receptor altered, involved in synaptic remodeling underlying learning and memory; two variants of α-integrin coded for
leonardo^{+}	Short-term and olfactory learning	Affects protein 14-3-3-, which is involved in intracellular signaling that activates and represses protein kinase C activity, activates tyrosine hydroxylase and tryptophan hydroxylase (enzymes involved in catecholamine and serotonin synthesis) serotonin synthesis)

Adapted from Dubnau and Tully (1998), Sokolowski (2001), Waddell and Quinn (2001).

dunce^{+} is expressed in the mushroom bodies in the brain of *D. melanogaster* (Figure 11.1). This was discovered because the mushroom bodies can be stained with an antibody to the *dunce*$^{+}$-encoded protein (Figure 11.1). The activity of the *dunce*$^{+}$ gene was identified by the **enhancer trap** method (O'Kane and Gehring 1987), a technique which involves placing a **reporter gene** (such as β-*galactosidase* which turns the fly's brain tissues blue when the substrate is added) into the *P* element under the control of a weak constitutive promoter. When this *P* element is brought in proximity to a tissue-specific enhancer after the *P* inserts into a chromosome, β-*galactosidase* expression will be regulated by the "native" enhancer in a tissue- and stage-specific pattern. Ideally, β-*galactosidase* will be expressed in a manner similar to the native gene. To determine which genes are expressed in the mushroom bodies, fly brains were screened and some 50 learning mutants were identified, including several alleles of *rutabaga*$^{+}$. Subsequently, mutations of *rutabaga*$^{+}$ were found to cause decreased expression of cAMP, and the rutabaga protein was identified as an adenylate cyclase (Han et al. 1992, Table 11.3).

Mushroom bodies are important for olfactory learning and memory. In *D. melanogaster* these structures are paired and consist of about 2500 neurons (Davis 1993, Heisenberg 1998, Figure 11.1). Mushroom bodies receive olfactory information from the antennal lobes. Mushroom bodies house part of the short-term memory for odors, are required for courtship conditioning memory, and are necessary for context generalization in visual learning, as well as regulating the transition from walking to rest (Zars 2000). By analyzing a *Drosophila* strain with *alpha-lobes-absent*, a mutation which causes flies to lack either the two vertical lobes of the mushroom body or two of the three median lobes which contain branches of the vertical lobe neurons, Pascual and Preat (2001) found that long-term memory requires the vertical lobes. Short-term memory was normal in flies lacking either vertical lobes or the two median lobes.

Learning requires other brain centers, including the antennal lobes, the central complex, and the lateral protocerebrum in insects (Davis 1993, Hansson and Anton 2000). During metamorphosis, the nervous system of holometabolous insects such as *Drosophila* changes significantly. A controversy has existed as to whether flies retain learned behavior after metamorphosis from larvae to adults. There is no evidence that larval conditioning induces a change in adult olfactory responses (Barron and Corbet 1999). This is not surprising, because larval sense organs undergo histolysis during the pupal stage and adult sense organs are formed *de novo* from imaginal discs. The mushroom bodies of the fly brain are extensively rewired during metamorphosis.

Drosophila carrying a mutant version of the *turnip*$^+$ gene have difficulty in olfactory discrimination, conditioning of leg position, and larval, visual, and reward learning (Table 11.3). The *turnip*$^+$ gene is located on the X chromosome and is associated with reduced protein kinase C activity (Choi et al. 1991). Specifically, *turnip* mutants are defective in phosphorylation of pp76, a membrane protein in head tissues. Protein phosphorylations have been implicated repeatedly in changes underlying learning and short-term memory.

Additional mutated genes, including *radish*, *amnesiac*, *cabbage*, *latheo*, and *linotte*, are involved in abnormal learning or memory of *D. melanogaster* (Table 11.3). For example, flies with the X-linked *radish* mutation initially learn in olfactory tests, but their subsequent memory decays rapidly at both early and late times after learning. The *radish* flies show normal locomotor activity and sensitivity to odor cues and electric-shock reinforcements used in the learning tests. Anesthesia-resistant memory, or consolidated memory, is strongly reduced in *D. melanogaster* with the *radish* phenotype (Folkers et al. 1993).

The *rutabaga*$^+$ gene codes for an adenylyl cyclase and is expressed in *Drosophila* mushroom bodies. This gene is involved in olfactory short-term memory (Zars et al. 2000). Likewise, *Volado*$^+$, which codes for an α-integrin that mediates cell adhesion and signal transduction, is expressed in mushroom body cells of *Drosophila* and mediates short-term memory in olfactory learning (Grotewiel et al. 1998). Integrins have diverse biological roles, including cell-cycle regulation, cell migration, and cell death (apoptosis), functioning as mediators of interactions between cells with the extracellular matrix. Integrins also can transduce information across cell membranes bidirectionally. Grotewiel et al. (1998) speculated that integrins might produce a rapid alteration in the structure and efficacy of a synapse, without the necessity for protein synthesis. Alternatively, integrins might function through ligand binding followed by intracellular signaling events, through Ca^{2+} mobilization, tyrosine kinase activation, or induction of protein kinase C.

The enlightenment obtained from the study of *Drosophila* learning mutants is providing an understanding of learning in higher organisms (Tully 1991a,b, 1996, Dubnau and Tully 1998, Sokolowski 2001, Waddell and Quinn 2001).

11.5.3. Functional Genomics of Odor Behavior in *Drosophila*

The ability to respond to odors is essential for survival and reproduction, allowing insects to select mates, find and choose food, and locate appropriate oviposition sites. A beginning has been made in understanding the complex genetic basis of odor behavior in insects using *D. melanogaster* as a model system (Field et al. 2000, Vosshall 2000, Anholt et al. 2001). Efforts also are being made to evaluate olfaction, learning, and memory in the honey bee in a comparative neurogenomics approach (Maleszka 2000).

Odors are received by olfactory receptors located on the antennae and the maxillary palps, which send their axons to the antennal lobes in insect brains. Each third antennal segment in *D. melanogaster* contains about 1300 olfactory receptor cells, and each maxillary palp carries 120 chemosensory neurons (Anholt et al. 2001). These neurons project to 43 glomeruli in the antennal lobe of the brain. From there, processed olfactory information is relayed to higher-order brain centers (the mushroom body and the lateral horn of the protocerebrum).

It is thought that there are fewer than 100 *types* of odor receptors in insects (Vosshall et al. 1999), perhaps as few as 50 or 60 (Vosshall et al. 2000). By contrast, mammals have more than 10,000 different receptor types. The molecular receptors in an olfactory system involve seven G-protein-coupled transmembrane proteins (Mombaerts 1999). Such proteins are found in the mouse and rat, where the number of such genes number approximately 1000 or nearly 1% of the genome, which certainly indicates that odor reception is an important component of the mammalian genome. Sequencing the *Drosophila* genome has allowed similar receptor proteins to be identified in an insect using a comparative genomics approach (Vosshall 2000, Clyne et al. 1999, Vosshall et al. 2000). The receptor genes were found by searching the *Drosophila* genome to identify sequences that might encode transmembrane domains. The receptors found consist of large multigene families (Clyne et al. 2000, Vosshall et al. 1999). Once an odor or pheromone has activated the olfactory receptors, it needs to be deactivated. Several enzymes have been found that appear to degrade odor stimulants, including esterases, oxidases, and glutathione transferases (Field et al. 2000).

D. melanogaster is able to recognize and discriminate between a large number of odorants (Vosshall 2001). Because there are as few as 50 or 60 *types* of receptors in insects, each olfactory sensory neuron responds to several odorants, but responds maximally to only one (Dryer 2000). Whereas the average olfactory receptor gene is expressed in 20 olfactory neurons, some receptor genes are expressed in only two to three neurons. Seven olfactory receptor genes are expressed solely in the maxillary palp (Vosshall et al. 2000).

The approximately 50 to 60 odorant receptor genes in insects encode a novel family of proteins with seven membrane-spanning domains; these genes are unrelated to vertebrate or nematode chemosensory receptors, suggesting the genes emerged in an independent manner during evolution (Dryer 2000, Vosshall 2001). Furthermore, the *Drosophila* genes are poorly grouped into subfamilies of similar sequences because they exhibit low levels of sequence similar to each other (Dryer 2000).

11.5.4. Learning in *Apis mellifera*

Mushroom bodies in the Hymenoptera are much larger than those in *Drosophila*, which may reflect the importance of the mushroom bodies for social behavior, learning, and memory in the honey bee (Rinderer 1986, Rybak and Menzel 1993, Meller and Davis 1996).

Many social Hymenoptera (ants, bees, wasps) have complex behaviors, including caring for their brood. Social bee species such as *Apis mellifera* feed, protect, and nurse larvae, store food, and respond to adverse environmental factors. They search for nectar and pollen at unpredictable sites; they learn the celestial and terrestrial cues that guide their foraging trips over long distances and allow them to find their nest sites once again. They learn how to respond to the changing position of the sun, to a pattern of polarized light during the day, and to landmarks. Associative learning is an essential component to foraging behavior and dance communication. Hive mates attending a dance performance learn the odor the dancing bee carries and seek out that same odor when they forage for food.

The complexity of bee behavior makes it an ideal organism to analyze to better understand learning, especially in response to odors (Hammer and Menzel 1995, Ray and Ferneyhough 1999, Galizia and Menzel 2000a). Associative olfactory learning in honey bees has several features similar to higher forms of learning in vertebrates (Grunbaum and Muller 1998).

11.5.5. Pheromones in Insects

Many insects use chemical cues as signals to find mates, and molecular genetic methods are now used to study various aspects of pheromone response behavior. For example, genes are being identified, characterized, and cloned that code for proteins involved in the synthesis of pheromones (a substance released by the body that causes a predictable reaction by another individual of the same species), the perception of semiochemicals (chemicals that influence insect interactions), and the processing of the signals (Krieger and Breer 1999, Mombaerts 1999, Tillman et al. 1999, Field et al. 2000).

Pheromone biosynthesis appears to use one or a few enzymes that convert the products of normal primary metabolism into compounds that act as pheromones (Tillman et al. 1999). For example, pheromones arise from isoprenoid biosynthesis, or by the transformation of amino acids or fatty acids. A number of genes encoding the enzymes involved in transforming metabolites into pheromones have been cloned and sequenced (Field et al. 2000). The production of pheromones by insects is regulated by three hormonal messengers: juvenile hormone III, ecdysteroids, and a neuropeptide called PBAN (pheromone biosynthesis activating neuropeptide).

Perception of volatile pheromones is mediated by olfactory organs (sensillae) located primarily on the antennae. Some receptor neurons on the antennae appear to respond to one particular chemical (specialist neurons), but others appear to respond to a number of compounds (generalist neurons). Pheromones often are perceived in combination with other chemicals, including plant volatiles.

The detection of pheromones and other chemicals by insects involves proteins (**odorant binding proteins, OBPs**) that carry the compounds from the surface of the antennal sensilla through the sensillum lymph to the G-protein-coupled receptors and the olfactory neurons (Prestwich 1996, Krieger and Breer 1999). The odorant binding proteins (which includes pheromone binding proteins) are small, soluble proteins that are concentrated in the sensillum lymph. Genes and cDNAs encoding OBPs of many insects have been cloned (Christophides et al. 2000). Analysis indicates that the binding proteins of unrelated species have low levels of amino acid sequence similarity, although they do have a conserved region with cysteines that may be important for function. It appears that there has been gene duplication and divergence of odorant binding protein genes, with moth proteins belonging to one branch and the proteins of other insects not closely related (Christophides et al. 2000).

Multielectrode recording of the *Manduca sexta* antennal lobe indicates that the relative timing of action potentials may convey information about odor concentration and mixture (Galizia and Menzel 2000b). Rapid progress in elucidating olfaction and gustation in insects promises to advance our understanding of how insects perceive chemical cues in their environment.

11.5.6. Neurobiochemistry of *Drosophila*

Molecular neurobiology is concerned about how the nervous system controls behavior at the molecular level (Glover and Hames 1989). What are the biochemical substrates of behavior? A molecular genetic approach using *Drosophila* is providing interesting answers for both insects and mammals. For example, a potassium channel gene family was cloned first from *Drosophila* and subsequently from humans and mice using probes from *Drosophila*. The *Shaker*$^+$, *Shal*$^+$, *Shab*$^+$, and *Shaw*$^+$ subfamilies of the K^+ channel gene family have been found in the Chordata, Arthropoda, and Mollusca, suggesting that the ancestral K^+ channel gene had already given rise to these subfamilies by the time of the Cambrian radiation (Salkoff et al. 1992).

A number of enzymes and receptors are involved in neurobiology, including receptors for neurotransmitters and hormones, ion channel proteins and associated signal transduction components, brain-specific protein kinases, enzymes for transmitter synthesis, neuropeptide processing enzymes, neuron-specific growth or survival trophic factors and their receptors, inhibitors of neuronal growth, glial-specific growth factors and their receptors, proteins associated with memory, neuronal cytoskeleton and axonal transport proteins, and others not listed here or yet to be identified. A major endeavor in molecular neurobiology involves establishing the primary structure of all the categories of proteins involved in nerve signal reception and transmission (Barnard 1989).

11.5.6.1. Electrical Signaling

The nervous system receives information about its internal and external environment, processes this information, and produces an appropriate response. The signaling of nerve cells depends on the electrical status of their outer membranes. Nerve cells maintain a potential difference across the membrane with the inside of the cell negative relative to the outside of the cell. The resting nerve cell also maintains concentration gradients of sodium, calcium, and potassium ions. Sodium and calcium ions are at a relatively high concentration outside the cell, while potassium ion levels are relatively high inside the cell. Signaling then involves a change in the resting membrane potential brought about by charge transfers carried by ionic fluxes through gated pores formed by transmembrane proteins called **channels**.

Ion channel proteins catalyze the transmembrane flow of ionic charge by forming narrow, hydrophilic pores through which ions can diffuse passively (Miller 1991). **Ion channels** must open or close rapidly in response to biological signals (= gating). Furthermore, the open pore is generally selective and will determine which ions will permeate and which will not (**ionic selectivity**). Thus, a specific channel will permit K^+ but not Na^+ to pass, even though these ions are not geometrically elaborate structures and are thus not recognized specifically by enzymes.

Stimuli from the environment are perceived by specialized nerve cells (sensory cells). Each type of sensory cell responds to a particular stimulus such as light, sound, touch, heat,

or chemicals such as pheromones. These sensory cells transform and amplify the energy provided by a stimulus into an electrical signal (= sensory transduction). **Sensory transduction** is probably due to an alteration in the ionic permeability of the sensory cell membrane, which causes a depolarization of the membrane of the sensory cell from its resting level. The amplitude and duration of this departure generally increases logarithmically with the intensity of the stimulus. This signal is local and is not transmitted along the nerve cell; however, it acts as a stimulus to the axon, and if the depolarization increases over a threshold level, the signal will trigger a change in **action potential** in the axon. Action potentials are all-or-nothing electrical impulses that propagate without distortion or attenuation along the entire length of an axon.

The generation and propagation of an action potential alters ionic conditions within the cell. When axonal membranes are depolarized, sodium channels open and allow sodium ions to flow down their gradient into the cell, producing the depolarizing phase of an action potential. Within milliseconds after the sodium channels are opened, they are inactivated, but at about the same time the membrane depolarization activates potassium channels, and the reciprocal K^+ flow repolarizes the cell and restores the membrane resting potential. During the course of an action potential, the sodium currents in one region of the axon membrane cause the depolarization and firing of an action potential in an adjacent region of the membrane so that the action potential is propagated along the full length of the axon.

The electrical signal is transmitted between cells at special sites called **synapses**, which occur between two nerve cells as well as between nerve cells and effectors such as muscle cells. The signal is relayed by a chemical neurotransmitter which is packaged in membrane-bound vesicles. When an action potential reaches the presynaptic terminal, the depolarization activates calcium channels in the presynaptic membrane and the subsequent influx of calcium ions leads to the release of neurotransmitter. The neurotransmitter diffuses to the postsynaptic cell and interacts with specific receptors on that cell surface. Receptors are activated in response to binding of the specific neurotransmitter molecules. Generally, the size and duration of a synaptic potential reflect the amount of transmitter released by the presynaptic terminal. By depolarizing the postsynaptic cell above the threshold, the synaptic potential triggers the generation of an action potential, which continues the signaling one step further along the neural pathway (Ganetzky and Wu 1989).

11.5.6.2. Neurotransmitters

Acetylcholine (ACh) is the major neurotransmitter in the central nervous system of *Drosophila* and other insects. Choline acetyltransferase (ChAT) is the biosynthetic enzyme, and acetylcholinesterase (AChE) is the degradative enzyme. AChE terminates synaptic transmission by rapidly hydrolyzing acetylcholine. Both enzymes are found in the *Drosophila* central nervous system, and their genes have been cloned (Ganetzky and Wu 1989, Fournier et al. 1989).

The acetylcholinesterase gene (*Ace*) from *Drosophila* is 34 kb long and is split into ten exons, with the splicing sites of the two last exons precisely conserved among *Drosophila* and vertebrate cholinesterases (Fournier et al. 1989). The deduced mature *Ace* transcript is 4.2 kb long. A gene for an acetylcholine receptor subunit has been identified and cloned, and the amino acid sequence of this AChR shares similarity with vertebrate sequences.

11.5.6.3. Ion Channels

Two types of ion channels, permeable to sodium (Na^+) or potassium (K^+) ions, are responsible for membrane electrical phenomena. The Na^+ and K^+ channels are encoded by multigene families. Genes that affect Na^+ channels have been cloned (Salkoff et al. 1987), including *nap*^{ts} (no action potential, temperature-sensitive) and *para* (paralytic). The *nap*^{ts} gene affects the level of Na^+ channel activity and, at high temperatures, causes paralysis associated with a loss of action potentials (Kernan et al. 1991). The mutation *para*^{ts} is a temperature-sensitive mutation that causes instantaneous paralysis of adults at 29°C and of larvae at 37°C and encodes Na^+ channels (Loughney et al. 1989). The *para* locus encodes a protein that shares regions of extensive amino acid similarity with the α subunit of vertebrate Na^+ channels. Mutations of several different genes (*Shaker*, *Shal*, *Shab*, and *Shaw*) alter K^+ currents (Covarrubias et al. 1991). One of the best-studied is *Shaker*. *D. melanogaster* carrying the *Shaker* allele exhibit aberrant behavior, shaking their legs when anesthetized with ether. The underlying mechanism for this phenotype has been determined using molecular analyses (Papazian et al. 1987).

Flies in which the *Shaker* gene is deleted still have K^+ currents, which suggests that K^+ channel proteins also are encoded by other genes. Butler et al. (1989) used a cDNA probe for *Shaker* and low-stringency hybridization of a cDNA library to isolate three additional family members, *Shab*$^+$, *Shaw*$^+$, and *Shal*$^+$. These genes are organized similarly to *Shaker*$^+$ in that only a single domain containing six presumed membrane-spanning segments is coded by each mRNA. These four genes define four K^+ channel subfamilies in *Drosophila*, and homologous genes isolated from vertebrates all appear to fall into one of these four subclasses.

Other K^+ channel mutants, including *ether-a-go-go* (Warmke et al. 1991) and a calcium-activated K^+ channel gene (*slo*) (Atkinson et al. 1991), have been isolated. Another neurotransmitter, γ-aminobutyric acid (GABA), is a major inhibitory agent in the insect nervous system. The synthesis of GABA is controlled by the enzyme glutamic acid decarboxylase (GAD) (Jackson et al. 1990).

11.5.7. Divergent Functions of *Est-6* and *Est-5* in Two *Drosophila* Species

Evolutionary changes in gene regulation can be important in macroevolutionary change and species divergence. One case study involves an analysis of the esterase 6 enzyme in *Drosophila melanogaster* and its homolog (esterase 5) in *D. pseudoobscura* (Brady and Richmond 1990). This gene influences behavior in *D. melanogaster* but has a very different function in *D. pseudoobscura.*

Esterase-6 (*Est-6*) in *D. melanogaster* influences male mating speed and rate of remating by females. Fast and slow variants of esterase 6 protein, as detected by electrophoresis, are produced in natural populations of *D. melanogaster*. More esterase 6 protein is produced in adult males than in females. The enzyme is highly concentrated in the anterior ejaculatory duct of males and is transferred to females during the first 2 to 3 min of the 20-min copulation. Enzyme activity in females can be detected up to 2 h after mating and influences the timing of remating by females. Males transfer a substance in the seminal fluid which is converted in the females' reproductive tract by esterase 6 into a pheromone that serves as an **antiaphrodisiac**. The antiaphrodisiac reduces the sexual attractiveness and receptivity of females, reducing the likelihood she will remate. Because the sperm from the most recent male takes precedence in fertilizing a female's eggs,

this behavior appears to encourage monogamy in *D. melanogaster* females (Richmond et al. 1986).

Est-6 also influences the rate of mating of males in *D. melanogaster*. Males with the slow variant of the protein require 10.2 min to achieve copulation with females, while males with the faster-moving protein require only 5.7 min. Once the *Est-6* gene was cloned, it was used as a probe to identify homologous genes in related species, which can provide clues to the evolution of behavior (Brady and Richmond 1990), and *Est-5* was isolated from *D. pseudoobscura*. Surprisingly, *Est-5* has a different function in *D. pseudoobscura*; it is expressed in the eyes and hemolymph. Despite these different patterns of expression, *Est-6* and *Est-5* have similar protein products, transcripts, and DNA sequences.

When *Est-5* from *D. pseudoobscura* was cloned into a *P* element and introduced into *D. melanogaster*, its activity and pattern of expression in *D. melanogaster* matched those of *D. pseudoobscura*, implying that regulatory elements had been conserved since the divergence of the two species 20 to 46 million years ago. Brady and Richmond (1990) speculated that the enzyme in the common ancestor of these two species had a more extensive expression pattern. After their divergence, regulatory mutations may have occurred that enhanced *Est-5* expression in the eyes of *D. pseudoobscura*, while mutations in *Est-6* led to increased expression in the male ejaculatory duct of *D. melanogaster*. Thus, the use of DNA sequence similarity to identify behavioral (and other) genes can lead to surprises.

11.5.8. Courtship Behavior in *Drosophila*

Mating behavior of *D. melanogaster* is stereotypical, with a fixed sequence of actions that are under genetic control. Courtship involves visual stimuli, acoustic signals, and pheromones (Hall 1994, Yamamoto et al. 1997, Goodwin 1999, Savarit et al. 1999, Greenspan and Ferveur 2000). Male courtship behavior involves six elements in the following fixed order: orienting → following → wing vibration → licking → attempting to copulate → copulation.

Sexual differentiation in *Drosophila*, described in Chapter 10, is controlled by a short cascade of regulatory genes, the expression of which determines all aspects of maleness and femaleness in the soma and the central nervous system. These genes also influence courtship behavior. Sexual behavior is irreversibly programmed during a critical period as a result of the activity, or inactivity, of the control gene *tra*$^+$. Male behavior is replaced by female behavior when *tra*$^+$ is expressed around the time of puparium formation (Arthur et al. 1998).

Other genes indirectly affect courtship behavior in *Drosophila*, including genes that involve general behavior (*yellow*$^+$, *inactive*$^+$, *couch potato*$^+$, *cuckold*$^+$, *minibrain*$^+$, *nerd*$^+$); visual behavior (*white*$^+$, *optomotor-blind*$^+$, *no-receptor-potential-A*$^+$); olfaction (*smellblind*$^+$); learning/memory genes (*dunce*$^+$, *rutabaga*$^+$, *amnesiac*$^+$, *Shaker*$^+$, *ether-a-go-go*$^+$); regulating periodicity of behavior (*period*$^+$); courtship song mutants (*cacophony*$^+$, *dissonance*$^+$, *croaker*$^+$, *fruitless*$^+$); and female receptivity (*spinster*$^+$) (Hall 1994).

The *fruitless* mutation is involved in both sex determination and courtship behavior and is active in the central nervous system (Hall 1994, Ryner et al. 1996, Goodwin 1999, Baker et al. 2001). Males with the *fruitless* mutation may court both females and males without copulating. Male flies expressing this gene are unable to bend their abdomens in the presence of females they are courting because they lack a male-specific muscle of Lawrence. Some *fruitless* mutations cause males to be homosexual (they court only males), while others cause males to be bisexual (they court both males and females) (Yamamoto et al. 1997).

The *fruitless*$^+$ gene is the first gene in a branch of the sex-determination hierarchy functioning specifically in the central nervous system, with mutants of this gene affecting nearly all aspects of male sexual behavior (Ryner et al. 1996, Villella et al. 1997, Goodwin et al. 2000). It is at least 140 kb long and produces a complex array of transcripts by using four promoters and alternative splicing; the male-specific transcripts are expressed in only a small fraction of the central nervous system (Goodwin et al. 2000).

Another mutation, *dissatisfaction*, is necessary for some aspects of sex-specific courtship behavior and neural differentiation in flies of both sexes. Mutant males are bisexual but, unlike *fruitless* males, attempt to copulate. Males with the *dissatisfaction* phenotype take longer to copulate with females. Females with the *dissatisfaction* phenotype are unreceptive to male advances during courtship and do not lay mature eggs (Goodwin 1999). Mating behavior of normal females involves the following sequence: stopping moving → offering the courting male a chance to lick the female's genitalia → allowing males to attempt copulation. A nonreceptive female leaves the courting male, and if the male pursues her, she may kick him. Nonreceptive virgin females persistently repel male approaches by lifting their abdomens up to block physically any contacts with males. Nonreceptive fertilized females lower their abdomens and extrude their ovipositors and eggs to repel males. Thus, female receptivity varies with age, diet, hormonal condition, and mating experience. The *spinster* mutation affects the sexual receptivity of females throughout their lives, and females with the *spinster* phenotype continuously leave, kick, or fend off courting males (Hall 1994, Suzuki et al. 1997).

Both *D. melanogaster* and *D. simulans* females produce contact pheromones,which consist of cuticular hydrocarbons that elicit wing displays by males (Ferveur 1997). These chemical signals have a low volatility, act at a very short distance (a few millimeters), and are perceived by contact rather than smell. Flies from a given strain, sex, and age produce a reproducible pattern of cuticular hydrocarbons, the biochemical pathway of which is under genetic control. The most important hydrocarbons involved are 7-tricosene and 7-pentacosene. One mutation, *Ngbo*, influences the ratios of 7-tricosene and 7-pentacosene in *D. simulans*. Another, *kete*, reduces the amount of 7-tricosene and all other linear hydrocarbons but does not affect the ratio (Ferveur and Jallon 1993). Flies homozygous for both *kete* and *Ngbo* have reduced viability and fertility, perhaps because they have very little 7-tricosene.

Experiments were conducted to eliminate genetically all known cuticular hydrocarbons in *D. melanogaster* in order to determine how mating behavior would be modified (Savarit et al. 1999). The results were surprising; contrary to expectation that *D. melanogaster* females lacking cuticular pheromones would induce no courtship by males, such females remained attractive. Additional analysis indicated that undetermined pheromone(s), probably also cuticular hydrocarbons, were present on both control and transgenic flies. Savarit et al. (1999) suggested that these newly discovered pheromones represent ancestral attractive substances in *D. melanogaster* and its sibling species.

11.5.9. Speciation Genes in *Drosophila*

Changes in sexual behavior can result in reproductive isolation between populations, leading to speciation. Studies of sexual behaviors in *Drosophila* species have led to different conclusions about the number of genes involved in speciation by this mechanism (Doi et al. 2001, Ting et al. 2001).

Analyses of the genetics of speciation usually involve crossing pairs of related species that do not normally mate, but will do so under laboratory conditions when given no choice. The progeny of such "interspecific" crosses then are examined to determine what phenotypes are related to their reproductive isolation. Reproductive isolation can be due to sterility of the hybrids (**postmating isolation**) or differences in mate preference (contributing to **premating isolation**). Study of postmating isolation mechanisms indicate that a number of genes (loci) are involved.

Premating isolation is thought to be a common cause of speciation in insects caused by a divergence in male sexual signals and female preferences. As a result, **assortative mating** occurs, with individuals preferring to mate with individuals who resemble themselves. Ting et al. (2001) studied the sexual isolation of two populations of *D. melanogaster* (M and Z forms). Z females strongly prefer Z males over M males; this preference is due to at least four loci on chromosome III that influence male behavior and at least three loci that influence female behavior, suggesting that premating isolation has a multigenic basis.

By contrast, Doi et al. (2001) used *D. ananassae* and its sibling species *D. pallidosa* to analyze sexual isolation. These species are almost completely isolated, but *ananassae* females no longer discriminate strongly against *pallidosa* males if the males are prevented from singing their songs (by removing their wings) or if females are prevented from hearing them (by removing their ears). This suggests that divergence in male song patterns and associated female preferences underlies this sexual isolation. The genetic basis of the preference of *ananassae* females for *ananassae* males appears to be a single dominant gene.

The divergence of acoustic signals alone appears to explain the isolation between the *ananassae* and *pallidosa* species, but the basis of mate choice in the M and Z forms of *D. melanogaster* appears to involve different signals, which probably are determined by multiple genes. The histories of these populations could explain the different isolation mechanisms. M and Z forms of *D. melanogaster* appear to have diverged in the same geographic area (sympatric speciation), but the *ananassae* and *pallidosa* species may have evolved while isolated geographically (parapatric species). Analyses of additional populations and species are required to resolve how many behavioral genes are involved in speciation (Butlin and Ritchie 2001).

11.6. Human Neurodegenerative Diseases and Addictions in *Drosophila*

Drosophila is perhaps unique among eukaryotes in the variety and level of sophistication that can be applied to understand its neurobiology and behavior. As a result, *Drosophila* is being studied to gain knowledge about various neurodegenerative diseases in humans (Mutsuddi and Nambu 1998, Andretic et al. 1999, Feany 2000, Fortini and Bonini 2000).

Modeling diseases in simple invertebrate systems is attractive because genetics can define cellular cascades mediating death of neurons in Parkinson's disease, the second most common neurodegenerative disorder in humans (Feany and Bender 2000). Transgenic *Drosophila* containing a mutant form of the human α-synuclein gene exhibit the essential features of Parkinson's disease in humans, making it possible to study the function of α-synuclein and determine the underlying pathogenic mechanisms in a genetically tractable animal.

The *spongecake* mutant of *Drosophila* shows degenerative changes similar to those seen in humans with Creutzfeld–Jakob disease, while the *eggroll* mutant produces changes

similar to those seen in humans with Tay–Sachs disease (Min and Benzer 1997). The *beta-amyloid protein precursor-like* (*Appl*) gene of *Drosophila* encodes a homolog of the human β-amyloid precursor protein which gives rise to β-amyloid, a major component of the plaques found in patients suffering from Alzheimer's disease (Luo et al. 1992). Another protein associated with Alzheimer's disease, presenilin, has been found in *Drosophila*, and studies suggest it also may be involved in the development of the pathology (Fortini and Bonini 2000). A *Drosophila* homolog was identified for the human gene for copper/zinc superoxide dismutase; mutants of this gene are implicated in Lou Gehrig's disease (McCabe 1995, Phillips et al. 1995).

A recessive mutant (*bubblegum*) in *D. melanogaster* exhibits adult neurodegeneration similar to that seen in the human disease adrenoleukodystrophy (ALD), otherwise known as the disease cured in the movie *Lorenzo's Oil* (Min and Benzer 1999). In ALD, high levels of very long chain fatty acids are produced that can be lowered by dietary treatment with a mixture of unsaturated fatty acids; feeding the ALD flies one of the components, glyceryl trioleate oil, blocked the accumulation of excess very long chain fatty acids and eliminated the development of pathology. Thus, *bubblegum* flies provide a model system for studying mechanisms of disease and screening drugs for treatment.

Drosophila may serve as a model organism to study the genetics of alcohol abuse and drug addiction in humans (Bellen 1998, Moore et al. 1998, Andretic et al. 1999, Wolf 1999, Bainton et al. 2000, Singh and Heberlein 2000). Alcohol addiction and many types of drug addictions appear to share common mechanisms (Bellen 1998, Moore et al. 1998). For example, the "dopamine hypothesis" suggests that addictive drugs may activate certain areas of the human brain, leading to an increase in dopamine neurotransmitter release (Bainton et al. 2000). Elevation of dopamine probably provides a sense of well-being, pleasure, or elation, resulting in a positive reinforcement. Dopamine is not the only neurotransmitter acting in alcohol abuse; glutamate, serotonin, and GABA also may be involved. Furthermore, four of the five circadian genes (*period*$^+$, *clock*$^+$, *cycle*$^+$, *doubletime*$^+$) in *D. melanogaster* influence the fly's responsiveness to cocaine and suggest a biochemical regulator of cocaine sensitization (Andretic et al. 1999).

Selection of *D. melanogaster* for resistance to ethanol was shown to be determined by multiple genetic components. Singh and Heberlein (2000) analyzed 23 mutant fly strains with different responses to ethanol, and the effects of acute ethanol exposure on *Drosophila* locomotor behaviors are "remarkably similar to those described for mammals." Thus, study of *Drosophila* "may pave the way for an in-depth study of the genes involved in acute and chronic effects of ethanol" (Bellen 1998). Bainton et al. (2000) showed that, as in mammals, dopaminergic pathways in *Drosophila* play a role in modulating specific behavioral responses to cocaine, nicotine, or ethanol.

Drosophila flies have been shown recently to sleep, and they may become a model for understanding sleep in other animals (Hendricks et al. 2000, Greenspan et al. 2001). Flies that are "resting" choose a preferred location and become immobile for periods of up to 157 min at a particular time in the circadian day, becoming relatively unresponsive to sensory stimuli. When rest is prevented, the flies tend to rest despite stimulation and exhibit a "rest rebound." Drugs that affect sleep in mammals alter rest in flies, suggesting conserved neural mechanisms.

"During sleep, an animal cannot forage for food, take care of its young, procreate or avoid the dangers of predation, indicating . . . sleep must serve an important function" (Greenspan et al. 2001). Sleep disorders in humans are common, but the genes underlying these disorders are unknown (Kolker and Turek 1999). Analysis of *Drosophila* behavior

at the molecular level offers promise of elucidating this evolutionarily important aspect of survival.

References Cited

Allada, R., P. Emery, J. S. Takahashi, and M. Rosbash (2001). Stopping time: The genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24**: 1091–1119.

Alcock, J. (1984). *Animal Behavior: An Evolutionary Approach.* Sinauer Assoc., Sunderland, MA.

Andretic, R., S. Chaney, and J. Hirsch (1999). Requirement of circadian genes for cocaine sensitization in *Drosophila*. *Science* **285**: 1066–1068.

Anholt, R. R. H., J. J. Fanara, G. M. Fedorowicz, I. Ganguly, N. H. Kulkarni, T. F. C. Mackay, and S. M. Rollmann (2001). Functional genomics of odor-guided behavior in *Drosophila melanogaster*. *Chem. Senses* **26**: 215–221.

Arathi, H. S., I. Burns, and M. Spivak (2000). Ethology of hygienic behaviour in the honey bee *Apis mellifera* L. (Hymenoptera: Apidae): Behavioural repertoire of hygienic bees. *Ethology* **106**: 365–379.

Arthur, Jr., B. I., J. M. Jallon, B. Caflisch, Y. Choffat, and R. R. Nothiger (1998). Sexual behaviour in *Drosophila* is irreversibly programmed during a critical period. *Curr. Biol.* **8**: 1187–1190.

Atkinson, N. S., G. A. Robertson, and B. Ganetzky (1991). A component of calcium-activated potassium channels encoded by the *Drosophila slo* locus. *Science* **253**: 551–555.

Baker, B. S., B. J. Taylor, and J. C. Hall (2001). Are complex behaviors specified by dedicated regulatory genes? Reasoning from *Drosophila. Cell* **105**: 13–24.

Bainton, R. J., L. T. Y. Tsai, C. M. Singh, M. S. Moore, W. S. Neckameyer, and U. Heberlein (2000). Dopamine modulates acute responses to cocaine, nicotine and ethanol in *Drosophila*. *Curr. Biol.* **10**: 187–194.

Bargiello, T. A., F. R. Jackson, and M. W. Young (1984). Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature* **312**: 752–754.

Barnard, E. A. (1989). Molecular neurobiology—an introduction, pp. 1–7. In: *Molecular Neurobiology*. D. M. Glover, and B. D. Hames, Eds. IRL Press, Oxford.

Barron, A. B., and S. A. Corbet (1999). Preimaginal conditioning in *Drosophila* revisited. *Animal Behav.* **58**: 621–628.

Barton Browne, L. (1993). Physiologically induced changes in resource-oriented behavior. *Annu. Rev. Entomol.* **38**: 1–25.

Beck, S. D. (1980). *Insect Photoperiodism.* 2nd ed. Academic Press, New York.

Bell, W. J. (1990). Searching behavior patterns in insects. *Annu. Rev. Entomol.* **35**: 447–467.

Bellen, H. J. (1998). The fruit fly: a model organism to study the genetics of alcohol abuse and addiction? *Cell* **93**: 909–912.

Bellen, H. J., S. Kooyer, D. D'Evelyn, and J. Pearlman (1992). The *Drosophila* Couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. *Genes Dev.* **6**: 2125–2136.

Benzer, S. (1973). Genetic dissection of behavior. *Sci. Am.* **229**: 24–27.

Beukeboom, L. W., and J. van den Assem (2001). Courtship and mating behaviour of interspecific *Nasonia* hybrids (Hymenoptera, Pteromalidae): a grandfather effect. *Behavior Genet.* **31**: 167–177.

Brady, J. P., and R. C. Richmond (1990). Molecular analysis of evolutionary changes in the expression of *Drosophila* esterases. *Proc. Natl. Acad. Sci. USA* **87**: 8217–8221.

Butler, A., A. Wei, K. Baker, and L. Salkoff (1989). A family of putative potassium channel genes in *Drosophila. Science* **243**: 943–947.

Butlin, R., and M. G. Ritchie (2001). Searching for speciation genes. *Nature* **412**: 31–32.

Chen, C., T. Malone, S. K. Beckendorf, and R. L. Davis (1987). At least two genes reside within a large intron of the *dunce* gene of *Drosophila*. *Nature* **329**: 721–724.

Choi, K. W., R. F. Smith, R. M. Buratowski, and W. G. Quinn (1991). Deficient protein kinase C activity in *turnip*, a *Drosophila* learning mutant. *J. Biol. Chemistry* **266**: 15999–16006.

Christophides, G. K., A. C. Mintzas, and K. Komitopoulou (2000). Organization, evolution and expression of a multigene family encoding putative members of the odorant binding protein family in the medfly *Ceratitis capitata. Insect Mol. Biol.* **9**: 185–195.

Clyne, P. J., C. G. Warr, M. R. Freeman, D. Lessing, J. Kim, and J. R. Carlson (1999). A novel family of divergent seven-transmembrane proteins: Candidate odorant receptors in *Drosophila*. *Neuron* **22**: 327–338.

Clyne, P. J., C. G. Warr, and J. R. Carlson (2000). Candidate taste receptors in *Drosophila*. *Science* **287**: 1830–1834.

Colvin, J., and A. G. Gatehouse (1993). Migration and genetic regulation of the pre-reproductive period in the cotton-bollworm moth, *Helicoverpa armigera. Heredity* **70**: 407–412.

Costa, R. A. A. Peixoto, G. Barbujani, and C. P. Kyriacou (1992). A latitudinal cline in a *Drosophila* clock gene. *Proc. R. Soc. London B* **250**: 43–49.

Covarrubias, M., A. Wei, and L. Salkoff (1991). *Shaker, Shal, Shab,* and *Shaw* express independent K^+ current systems. *Neuron* **7**: 763–773.

Davis, R. L. (1993). Mushroom bodies and *Drosophila* learning. *Neuron* **11**: 1–14.

Davis, R. L. (1996). Physiology and biochemistry of *Drosophila* learning mutants. *Physiol. Rev.* **76**: 299–317.

Davis, R. L., and B. Dauwalder (1991). The *Drosophila dunce* locus: learning and memory genes in the fly. *Trends Genet.* **7**: 224–229.

deBelle, J. S., and M. B. Sokolowski (1987). Heredity of rover/sitter: alternative foraging strategies of *Drosophila melanogaster* larvae. *Heredity* **59**: 73–83.

Delgado, R., R. Latorre, and P. Labarca (1992). K^+-channel blockers restore synaptic plasticity in the neuromuscular junction of *dunce*, a *Drosophila* learning and memory mutant. *Proc. R. Soc. Lond. B* **250**: 181–185.

Dingle, H., Ed. (1978). *Evolution of Insect Migration and Diapause.* Springer-Verlag, New York.

Dingle, H., and J. P. Hegmann, Eds. (1982). *Evolution and Genetics of Life Histories.* Springer-Verlag, New York.

Doerge, R. W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev. Genet.* **3**: 43–52.

Doi, M., M. Matsuda, M. Tomaru, H. Matsubayasi, and Y. Oguma (2001). A locus for female discrimination behavior causing sexual isolation in *Drosophila. Proc. Natl. Acad. Sci. USA* **98**: 6714–6719.

Dryer, L. (2000). Evolution of odorant receptors. *BioEssays* **22**: 803–810.

Dubnau, J., and T. Tully (1998). Gene discovery in *Drosophila*: new insights for learning and memory. *Annu. Rev. Neurosci.* **21**: 407–444.

Dunlap, J. C. (1998). Common threads in eukaryotic circadian systems. *Curr. Opin. Genet. Dev.* **8**: 400–406.

Dunlap, J. C. (1999). Molecular bases for circadian clocks. *Cell* **96**: 271–290.

Egan, E. S., T. M. Franklin, M. J. Hilderbrand-Chae, G. P. McNeil, M. A. Roberts, A. J. Schroeder, X. Zhang, and F. R. Jackson (1999). An extraretinally expressed insect cryptochrome with similarity to the blue light photoreceptors of mammals and plants. *J. Neurosci.* **19**: 366–367.

Ehrman, L., and P. A. Parsons (1973). *The Genetics of Behavior*. Sinauer Assoc., Sunderland, MA.

Ehrman, L., and P. A. Parsons (1981). *Behavior Genetics and Evolution*. McGraw-Hill, New York.

Emery, P., R. Stanewsky, C. Helfric-Forster, M. Emery-Le, J. C. Hall, and M. Rosbash (2000). *Drosophila* CRY is a deep brain circadian photoreceptor. *Neuron* **26**: 493–504.

Falconer, D. S. (1989). *Introduction to Quantitative Genetics*. 3rd ed. Longman, New York.

Feany, M. B. (2000). Studying human neurodegenerative diseases in flies and worms. *J. Neuropathol. Exp. Neurol.* **59**: 847–856.

Feany, M. B., and W. W. Bender (2000). A *Drosophila* model of Parkinson's disease. *Nature* **404**: 394–398.

Ferveur, J. F. (1997). The pheromonal role of cuticular hydrocarbons in *Drosophila melanogaster*. *BioEssays* **19**: 353–358.

Ferveur, J. F., and J. M. Jallon (1993). Genetic control of pheromones in *Drosophila simulans*. II. *kete* a locus on the X chromosome. *Genetics* **133**: 561–567.

Field, L. M., J. A. Pickett, and L. J. Wadhams (2000). Molecular studies in insect olfaction. *Insect Mol. Biol.* **9**: 545–551.

Folkers, E., P. Drain, and W. G. Quinn (1993). *radish*, a *Drosophila* mutant deficient in consolidated memory. *Proc. Natl. Acad. Sci. USA* **90**: 8123–8127.

Fortini, M. E., and N. M. Bonini (2000). Modeling human neurodegenerative diseases in *Drosophila*. On a wing and a prayer. *Trends Genet.* **16**: 161–167.

Foster, S. P., and M. O. Harris (1997). Behavioral manipulation methods for insect pest-management. *Annu. Rev. Entomol.* **42**: 123–146.

Fournier, D., F. Karch, J. M. Bride, L. M. C. Hall, J. B. Berge, and P. Spierer (1989). *Drosophila melanogaster* acetylcholinesterase gene structure, evolution and mutations. *J. Mol. Biol.* **210**: 15–22.

Gailey, D. A., A. Villella, and T. Tully (1991). Reassessment of the effect of biological rhythm mutations on learning in *Drosophila melanogaster*. *J. Comp. Physiol. A* **169**: 685–697.

Galizia, C. G., and R. Menzel (2000a). Odour perception in honeybees: coding information in glomerular patterns. *Curr. Opin. Neurobiol.* **10**: 504–510.

Galizia, C. G., and R. Menzel (2000b). Probing the olfactory code. *Nature Neurosci.* **3**: 853–854.

Ganetzky, B., and C. F. Wu (1989). Molecular approaches to neurophysiology in *Drosophila*, pp. 9–61. In: *Molecular Neurobiology.* D. M. Glover, and B. D. Hames, Eds. IRL Press, Oxford.

Gatehouse, A. G. (1989). Genes, environment and insect flight, pp. 116–138. In: *Insect Flight*. G. J. Goldsworthy, and C. Wheeler, Eds. CRC Press, Boca Raton, FL.

Gatehouse, A. G. (1997). Behavior and ecological genetics of wind-borne migration by insects. *Annu. Rev. Entomol.* **42**: 475–502.

Giebultowicz, J. M. (2000). Molecular mechanism and cellular distribution of insect circadian clocks. *Annu. Rev. Entomol.* **45**: 769–793.

Glover, D. M., and B. D. Hames, Eds. (1989). *Molecular Neurobiology*. IRL Press, Oxford, UK.

Goodwin, S.F. (1999). Molecular neurogenetics of sexual differentiation and behaviour. *Curr. Opin. Neurobiol.* **9**: 759–765.

Goodwin, S. F., B. J. Taylor, A. Villella, M. Foss, L. C. Ryner, B. S. Baker, and J. C. Hall (2000). Aberrant splicing and altered spatial expression patterns in *fruitless* mutants of *Drosophila melanogaster*. *Genetics* **154**: 725–745.

Gould, F. (1986). Genetic constraints on the evolution of cannibalism in *Heliothis virescens*, pp. 55–62. In: *Evolutionary Genetics of Invertebrate Behavior, Progress and Prospects.* M. D. Huettel, Ed. Plenum, New York.

Greenspan, R. J., and J. F. Ferveur (2000). Courtship in *Drosophila*. *Annu. Rev. Genet.* **34**: 205–232.

Greenspan, R. J., G. Tononi, C. Cirelli, and P. J. Shaw (2001). Sleep and the fruit fly. *Trends Neurosci.* **24**: 142–145.

Grossfield, J. (1975). Behavioral mutants of *Drosophila*, pp. 679–702. In: *Handbook of Genetics.* Vol. 3. *Invertebrates of Genetic Interest*. R. C. King, Ed. Plenum, New York.

Grotewiel, M. S., C. D. O. Beck, K. H. Wu, X. R. Zhu, and R. L. Davis (1998). Integrin-mediated short-term memory in *Drosophila. Nature* **391**: 455–460.

Grunbaum, L., and U. Muller (1998). Induction of a specific olfactory memory leads to a long-lasting activation of protein kinase C in the antennal lobe of the honeybee. *J. Neurosci.* **18**: 4384–4392.

Guantieri, V., A. Pepe, M. Zordan, C. P. Kyriacou, R. Costa, and A. M. Tamburro (1999). Different *period* gene repeats take "turns" at fine-tuning the circadian clock. *Proc. R. Soc. London B* **266**: 2283–2288.

Hall, J. C. (1984). Complex brain, and behavioral functions disrupted by mutations in *Drosophila*. *Dev. Genet.* **4**: 355–532.

Hall, J. C. (1985). Genetic analysis of behavior in insects, pp. 287–373. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology*. Vol. 9. G. A. Kerkut, and L. I. Gilbert, Eds. Pergamon, New York.

Hall, J. C. (1994). The mating of a fly. *Science* **264**: 1702–1714.

Hall, J. C. (1995). Tripping along the trail to the molecular mechanisms of biological clocks. *Trends Neurosci.* **5**: 230–240.

Hall, J. C. (1998a). Genetics of biological rhythms in *Drosophila*. *Adv. Genet.* **38**: 135–184.

Hall, J. C. (1998b). Molecular neurogenetics of biological rhythms. *J. Neurogenet.* **12**: 115–181.

Hammer, M., and R. Menzel (1995). Learning and memory in the honeybee. *J. Neurosci.* **15**: 1617–1630.

Han, P. L., L. R. Levin, R. R. Reed, and R. L. Davis (1992). Preferential expression of the *Drosophila rutabaga* gene in mushroom bodies, neural centers for learning in insects. *Neuron* **9**: 619–627.

Hansson, B. S., and S. Anton (2000). Function and morphology of the antennal lobe: new developments. *Annu. Rev. Entomol.* **45**: 203–231.

Hardin, P. E., J. C. Hall, and M. Rosbash (1992). Circadian oscillations in *period* gene mRNA levels are transcriptionally regulated. *Proc. Natl. Acad. Sci. USA* **89**: 11711–11715.

Harmer, S. L., S. Panda, and S. A. Kay (2001). Molecular bases of circadian rhythms. *Annu. Rev. Cell Dev. Biol.* **17**: 215–253.

Hay, D. A. (1985). *Essentials of Behaviour-Genetics*. Blackwell, Palo Alto, CA.

Heisenberg, M. (1997). Genetic approach to neuroethology. *BioEssays* **19**: 1065–1073.

Heisenberg, M. (1998). What do the mushroom bodies do for the insect brain? An introduction. *Learning Memory* **5**: 1–10.

Hendricks, J. C., S. M. Finn, K. A. Panckeri, J. Chavkin, J. A. Williams, A. Sehgal, and A. I. Pack (2000). Rest in *Drosophila* is a sleep-like state. *Neuron* **24**: 129–138.

Holman, G. M., R. J. Nachman, and M. S. Wright (1990). Insect neuropeptides. *Annu. Rev. Entomol.* **35**: 201–217.

Homberg, U., T. A. Christensen, and J. G. Hildebrand (1989). Structure and function of the deutocerebrum in insects. *Annu. Rev. Entomol.* **34**: 477–501.

Hotta, Y., and S. Benzer (1972). Mapping of behaviour in *Drosophila* mosaics. *Nature* **240**: 527–535.

Huber, F., T. E. Moore, and W. Loher, Eds. (1989). *Cricket Behavior and Neurobiology.* Cornell Univ. Press, Ithaca, NY.

Huettel, M. D., Ed. (1986). *Evolutionary Genetics of Invertebrate Behavior. Progress and Prospects.* Plenum, New York.

Huettel, M. D., and G. L. Bush (1972). The genetics of host selection and its bearing on sympatric speciation in *Procecidochares* (Diptera: Tephritidae). *Entomol. Exp. Appl.* **15**: 465.

Hunt, G. J., E. Guzman-Novoa, M. K. Fondrk, and R. E. Page, Jr. (1998). Quantitative trait loci for honey bee stinging behavior and body size. *Genetics* **1478**: 1203–1213.

Ishida, N., M. Kaneko, and R. Allada (1999). Biological clocks. *Proc. Natl. Acad. Sci. USA* **96**: 8819–8820.

Iwasaki, K., and J. H. Thomas (1997). Genetics in rhythm. *Trends Genet.* **13**: 111–115.

Jackson, F. R., L. M. Newby, and S. J. Kulkarni (1990). *Drosophila* GABAergic systems: sequence and expression of glutamic acid decarboxylase. *J. Neurochem.* **54**: 1068–1078.

Jaenike, J. (1986). Genetic complexity of host selection behavior in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **83**: 2148–2151.

Kalmring, K., and N. Elsner, Eds. (1985). *Acoustic and Vibrational Communication in Insects.* Paul Parey, Hamburg.

Kernan, M. J., M. I. Kuroda, R. Kreber, B. S. Baker, and B. Ganetzky (1991). nap^{ts}, a mutation affecting sodium channel activity in *Drosophila*, is an allele of *mle*, a regulator of X chromosome transcription. *Cell* **66**: 949–959.

Klun, J. A., and M. D. Huettel (1988). Genetic regulation of sex pheromone production and response: interaction of sympatric pheromonal types of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). *J. Chem. Ecol.* **13**: 2047–2061.

Klun, J. A., and S. Maini (1979). Genetic basis of an insect chemical communication system. *Environ. Entomol.* **8**: 423–426.

Kolker, D. E., and F. W. Turek (1999). The search for circadian clock and sleep genes. *J. Psychopharmacol*. 13(4) *Suppl.* **1**: S5–S9.

Krieger, J., and H. Breer (1999). Olfactory reception in invertebrates. *Science* **286**: 720–728.

Kyriacou, C. P. (1990). The molecular ethology of the *period* gene in *Drosophila*. *Behavior Genet.* **20**: 191–211.

Kyriacou, C. P. (1993). Molecular genetics of complex behavior: biological rhythms in *Drosophila*, pp. 329–356. In: *Molecular Approaches to Fundamental and Applied Entomology*. J. Oakeshott, and M. J. Whitten, Eds. Springer-Verlag, New York.

Kyriacou, C. P., and J. C. Hall (1989). Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc. Natl. Acad. Sci. USA* **77**: 6729–6733.

Lakin-Thomas, P. (2000). Circadian rhythms. New functions for old clock genes? *Trends Genet*. **16**: 135–142.

Liu, X., L. J. Zwiebel, D. Hinton, S. Benzer, J. C. Hall, and M. Rosbash (1992). The *period* gene encodes a predominantly nuclear protein in adult *Drosophila*. *J. Neurosci.* **12**: 2735–2744.

Lofstedt, C. (1990). Population variation and genetic control of pheromone communication systems in moths. *Entomol. Exp. Appl.* **54**: 199–218.

Lofstedt, C., B. S. Hansson, W. Roelofs, and B. O. Bengtsson (1989). No linkage between genes controlling female pheromone production and male pheromone response in the European corn borer, *Ostrinia nubilalis* Hubner (Lepidoptera: Pyralidae). *Genetics* **123**: 553–556.

Loudon, A. S. I., A. G. Semikhodskii, and S. K. Crosthwaite (2000). A brief history of circadian time. *Trends Genet.* **16**: 477–481.

Loughney, K., R. Kreber, and B. Ganetzky (1989). Molecular analysis of the *para* locus, a sodium channel gene in *Drosophila*. *Cell* **58**: 1143–1154.

Luo, L., T. Tully, and K. White (1992). Human amyloid precursor protein ameliorates behavioral deficit of flies deleted for *Appl* gene. *Neuron* **9**: 595–605.

Maleszka, R. (2000). Molecules to behaviour in the honeybee—-the emergence of comparative neurogenomics. *Trends Neurosci.* **23**: 513–514.

Masler, E. P., T. J. Kelly, and J. J. Menn (1993). Insect neuropeptides: discovery and application in insect management. *Arch. Insect Biochem. Physiol.* **22**: 87–111.

Masterman, R., R. Ross, K. Mesce, and M. Spivak (2001). Olfactory and behavioral response thresholds to odors of diseased brood differ between hygienic and non-hygienic honey bees (*Apis mellifera* L.). *J. Comp. Physiol. A* **187**: 441–452.

Matthews, R. W., and J. R. Matthews (1978). *Insect Behavior*. Wiley Interscience, New York.

McCabe, E. R. B. (1995). Modeling Lou Gehrig's disease in the fruit fly. *Proc. Natl. Acad. Sci. USA* **92**: 8533–8534.

McGuire, T. R. (1984). Learning in three species of Diptera: the blow fly, *Phormia regina,* the fruit fly, *Drosophila melanogaster*, and the house fly, *Musca domestica. Behav. Genet.* **14**: 479–526.

McGuire, T. R., and J. Hirsch (1977). Behavior-genetic analysis of *Phormia regina*: conditioning, reliable individual differences, and selection. *Proc. Natl. Acad. Sci. USA* **74**: 5193–5197.

McGuire, T. R., and T. Tully (1987). Characterization of genes involved with classical conditioning that produce differences between bidirectionally selected strains of the blow fly *Phormia regina*. *Behav. Genet.* **17**: 97–107.

Meller, V. H., and R. L. Davis (1996). Biochemistry of insect learning: lessons from bees and flies. *Insect. Biochem. Mol. Biol.* **26**: 327–335.

Menzel, R. (1999). Memory dynamics in the honey bee. *J. Comp. Physiol. A* **185**: 323–340.

Merrow, M., and T. Roenneberg (2001). The circadian cycle: is the whole greater than the sum of its parts? *Trends Genet.* **17**: 4–7.

Miller, C. (1991). **1990**: Annus mirabilis of potassium channels. *Science* **251**: 1092–1096.

Min, K. T., and S. Benzer (1997). *Spongecake* and *eggroll*: two hereditary diseases in *Drosophila* resemble patterns of human brain degeneration. *Curr. Biol.* **7**: 885–888.

Min, K. T., and S. Benzer (1999). Preventing neurodegeneration in the *Drosophila* mutant *bubblegum*. *Science* **284**: 1985–1988.

Mollema, C. (1991). Heritability estimates of host selection behaviour by the parasitoid *Asobara tabida*. *Neth. J. Zool.* **41**: 174–183.

Mombaerts, P. (1999). Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**: 707–711.

Moore, M. S., J. DeZazzo, A. Y. Luk, T. Tully, C. M. Singh, and U. Heberlein (1998). Ethanol intoxication in *Drosophila*: genetic and pharmacological evidence for regulation by the cAMP signaling pathway. *Cell* **93**: 997–1007.

Moretto, G., L. S. Goncalves, and D. De Jong (1993). Heritability of africanized and european honey bee defensive behavior against the mite *Varroa jacobsoni*. *Rev. Brasil. Genet.* **16**: 71–77.

Moritz, R. F. A. (1988). A reevaluation of the two-locus model for hygienic behavior in honey bees (*Apis mellifera* L.). *J. Hered.* **79**: 257–262.

Moritz, R. F. A. (1994). Selection for varroatosis resistance in honey bees. *Parasitol. Today* **10**: 236–238.

Mutsuddi, M., and J. R. Nambu (1998). Neural disease: *Drosophila* degenerates for a good cause. *Curr. Biol.* **8**: R809–R811.

Nassel, D. R. (1993). Neuropeptides in the insect brain: a review. *Cell Tissue Res.* **273**: 1–29.

O'Kane, C. J., and W. J. Gehring (1987). Detection *in situ* of genomic regulatory elements in *Drosophila. Proc. Natl. Acad. Sci. USA* **84**: 9123–9127.

Osborne, K. A., A. Robichon, E. Burgess, S. Butland, R. A. Shaw, A. Coulthard, H. S. Pereira, R. J. Greenspan, and M. B. Sokolowski (1997). Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* **277**: 834–836.

Palmer, J. O., and H. Dingle (1989). Responses to selection on flight behavior in a migratory population of milkweed bug (*Oncopeltus fasciatus*). *Evolution* **43**: 1805–1808.

Papaj, D. R., and R. J. Prokopy (1989). Ecological and evolutionary aspects of learning in phytophagous insects. *Annu. Rev. Entomol.* **34**: 315–350.

Papazian, D. M., T. L. Schwarz, B. L. Tempel, Y. N. Jan, and L. Y. Jan (1987). Cloning of genomic and complementary DNA from *Shaker*, a putative potassium channel gene from *Drosophila*. *Science* **237**: 749–753.

Pascual, A., and T. Preat (2001). Localization of long-term memory within the *Drosophila* mushroom body. *Science* **294**: 1115–1117.

Pavlidis, P., M. Ramaswami, and M. A. Tanouye (1994). The *Drosophila easily shocked* gene: a mutation in a phospholipid synthetic pathway causes seizure, neuronal failure, and paralysis. *Cell* **79**: 23–33.

Pereira, H. S., D. E. MacDonald, A. J. Hilliker, and M. B. Sokolowski (1995). *Chaser (Csr)*, a new gene affecting larval foraging behavior in *Drosophila melanogaster*. *Genetics* **141**: 263–270.

Phillips, J. P., J. A. Tainr, E. D. Getzoff, G. L. Boulianne, K. Kirby, and A. J. Hilliker (1995). Subunit-destabilizing mutations in *Drosophila* copper/zinc superoxide dismutase: neuropathology and a model of dimer dysequilibrium. *Proc. Natl. Acad. Sci. USA* **92**: 8574–8578.

Plomin, R. (1990). The role of inheritance in behavior. *Science* **248**: 183–188.

Prestwich, G. D. (1996). Proteins that smell: pheromone recognition and signal transduction. *Bioorg. Medic. Chem.* **4**: 505–513.

Qiu, Y., C. N. Chen, T. Malone, L. Richter, S. K. Beckendorf, and R. L. Davis (1991). Characterization of the memory gene *dunce* of *Drosophila melanogaster*. *J. Mol. Biol.* **222**: 553–565.

Raina, A. K., and J. J. Menn (1993). Pheromone biosynthesis activating neuropeptide: from discovery to current status. *Arch. Insect Biochem. Physiol.* **22**: 141–151.

Ray, S., and B. Ferneyhough (1999). Behavioral development and olfactory learning in the honeybee (*Apis mellifera*). *Dev. Psychobiol.* **34**: 21–27.

Renou, M., and A. Guerrero (2000). Insect parapheromones in olfaction research and semiochemical-based pest control strategies. *Annu. Rev. Entomol.* **45**: 605–630.

Richmond, R. C., S. D. Mane, and L. Tompkins (1986). The behavioral effects of a carboxylesterase in *Drosophila*, pp. 223–236. In: *Evolutionary Genetics of Invertebrate Behavior, Progress and Prospects.* M. D. Huettel, Ed. Plenum, New York.

Rinderer, T. E., Ed. (1986). *Bee Genetics and Breeding.* Academic Press, New York.

Rinderer, T. E., and A. M. Collins (1986). Behavioral genetics, pp. 155–176. In: *Bee Genetics and Breeding.* T. E. Rinderer, Ed. Academic Press, New York.

R'Kha, S., P. Capy, and J. R. David (1991). Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. *Proc. Natl. Acad. Sci. USA* **88**: 1835–1839.

Rothenbuhler, W. C. (1964). Behavior genetics of nest cleaning in honey bees. IV. Responses of F_1 and backcross generations to disease-killed brood. *Amer. Zool.* **4**: 111–123.

Rybak, J., and R. Menzel (1993). Anatomy of the mushroom bodies in the honey bee brain: the neuronal connections of the alpha-lobe. *J. Comp. Neurol.* **334**: 444–465.

Ryner, L. C., S. F. Goodwin, D. H. Castrillon, A. Anand, A. Villella, B. S. Baker, J. C. Hall, B. J. Taylor, and S. A. Wasserman (1996). Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell* **87**: 1079–1089.

Salkoff, L., A. Butler, A. Wei, N. Scavarda, K. Giffen, C. Ifune, R. Goodman, and G. Mandel (1987). Genomic organization and deduced amino acid sequence of a putative sodium channel gene in *Drosophila*. *Science* **237**: 744–753.

Salkoff, L., K. Baker, A. Butler, M. Covarrubias, M. D. Pak, and A. Wei (1992). An essential "set" of K^+ channels conserved in flies, mice and humans. *Trends Neurosci.* **15**: 161–166.

Sancar, A. (2000). CRYPTOCHROME: The second photoactive pigment in the eye and its role in circadian photoreception. *Annu. Rev. Biochem.* **69**: 31–67.

Sato, Y., M. Oguchi, N. Menjo, K. Imai, H. Saito, M. Ideda, M. Isobe, and O. Amashita (1993). Precursor polyprotein for multiple neuropeptides secreted from the suboesophageal ganglion of the silkworm *Bombyx mori*: characterization of the cDNA encoding the diapause hormone precursor and identification of additional peptides. *Proc. Natl. Acad. Sci. USA* **90**: 3251–3255.

Saunders, D. S. (1990). The circadian basis of ovarian diapause regulation in *Drosophila melanogaster*: is the *period* gene causally involved in photoperiodic time measurement? *J. Biol. Rhythms* **5**: 315–331.

Savarit, F., G. Sureau, M. Cobb, and J. F. Ferveur (1999). Genetic elimination of known pheromones reveals the fundamental bases of mating and isolation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **96**: 9015–9020.

Scharrer, B. (1987). Insects as models in neuroendocrine research. *Annu. Rev. Entomol.* **32**: 1–16.

Sehgal, A., J. Price, and M. W. Young (1992). Ontogeny of a biological clock in *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* **89**: 1423–1427.

Shaver, S. A., C. J. Varnam, A. J. Hilliker, and M. B. Sokolowski (1998). The foraging gene affects adult but not larval olfactory-related behavior in *Drosophila melanogaster. Behav. Brain Res.* **95**: 23–29.

Singh, C. M., and U. Heberlein (2000). Genetic control of acute ethanol-induced behaviors in *Drosophila*. Alcoholism: *Clin. Exp. Res.* **24**: 1127–1136.

Sokolowski, M. B. (2001). *Drosophila:* genetics meets behaviour. *Nat. Rev. Genet.* **2**: 879–890.

Sokolowski, M. B., H. S. Pereira, and K. Hughes (1997). Evolution of foraging behavior in *Drosophila* by density-dependent selection. *Proc. Natl. Acad. Sci. USA* **94**: 7373–7377.

Spivak, M., and M. Gilliam (1993). Facultative expression of hygienic behaviour of honey bees in relation to disease resistance. *J. Apic. Res.* **32**: 147–157.

Spivak, M., and M. Gilliam (1998). Hygienic behaviour of honey bees and its application for control of brood diseases and varroa. II. Studies on hygienic behaviour since the Rothenbuhler era. *Bee World* **79**: 169–186.

Strausfeld, N. J. (1976). *Atlas of an Insect Brain.* Springer-Verlag, Berlin.

Strausfeld, N. J., L. Hansen, Y. Li, R. S. Gomez, and K. Ito (1998). Evolution, discovery, and interpretations of arthropod mushroom bodies. *Learning Memory* **5**: 11–37.

Suzuki, K., N. Juni, and D. Yamamoto (1997). Enhanced mate refusal in female *Drosophila* induced by a mutation in the *spinster* locus. *Appl. Entomol. Zool.* **32**: 235–243.

Takahashi, J. S. (1992). Circadian clock genes are ticking. *Science* **258**: 238–239.

Takeda, M., and S. D. Skopik (1997). Photoperiodic time measurement and related physiological mechanisms in insects and mites. *Annu. Rev. Entomol.* **42**: 323–349.

Tauber, M. J., C. A. Tauber, and S. Masaki (1986). *Seasonal Adaptations of Insects.* Oxford Univ. Press, Oxford, UK.

Tazima, Y., H. Doira, and H. Akai (1975). The domesticated silkmoth, pp. 63–124. In: *Handbook of Genetics.* Vol. 3. *Invertebrates of Genetic Interest*. R. C. King, Ed. Plenum, New York.

Tillman, J. A., S. J. Seybold, R.A. Jurenka, and G. J. Blomquist (1999). Insect pheromones—an overview of biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.* **29**: 481–514.

Ting, C. T., A. Takahashi, and C.I. Wu (2001). Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. *Proc. Natl. Acad. Sci.* **98**: 6709–6713.

Trpis, M., and W. Hausermann (1978). Genetics of house-entering behaviour in East African populations of *Aedes aegypti* (L.) (Diptera: Culicidae) and its relevance to speciation. *Bull. Entomol. Res.* **68**: 521–532.

Tully, T. (1991a). Genetic dissection of learning and memory in *Drosophila melanogaster*, pp. 29–66. In: *Neurobiology of Learning, Emotion and Affect*. J. Madden IV, Ed. Raven Press, New York.

Tully, T. (1991b). Physiology of mutations affecting learning and memory in *Drosophila*—the missing link between gene product and behavior. *Trends Neurosci.* **14**: 163–164.

Tully, T. (1996). Discovery of genes involved with learning and memory: an experimental synthesis of Hirschian and Benzerian perspectives. *Proc. Natl. Acad. Sci. USA* **93**: 13460–13467.

Via, S. (1990). Ecological genetics and host adaptation in herbivorous insects: the experimental study of evolution in natural and agricultural systems. *Annu. Rev. Entomol.* **35**: 421–426.

Villella, A., D. A. Gailey, B. Berwald, S. Ohshima, P. T. Barnes, and J. C. Hall (1997). Extended reproductive roles of the *fruitless* gene in *Drosophila melanogaster* revealed by behavioral analysis of new *fru* mutants. Genetics **147**: 1107–1130.

Vosshall, L. B. (2000). Olfaction in *Drosophila*. *Curr. Opin. Neurobiol.* **10**: 498–503.

Vosshall, L. B. (2001). The molecular logic of olfaction in *Drosophila*. *Chem. Senses* **26**: 207–213.

Vosshall, L. B. H. Amrein, P. S. Morozov, A. Rzhetsky, and R. Axel (1999). A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**: 725–736.

Vosshall, L. B., A. M. Wong, and R. Axel (2000). An olfactory sensory map in the fly brain. *Cell* **102**: 147–160.

Waddell, S., and W. G. Quinn (2001). Flies, genes and learning. *Annu. Rev. Neurosci.* **24**: 1283–1309.

Wager-Smith, K., and S. A. Kay (2000). Circadian rhythm genetics: from flies to mice to humans. *Nat. Genet.* **26**: 23–27.

Warmke, J., R. Drysdale, and B. Ganetzky (1991). A distinct potassium channel polypeptide encoded by the *Drosophila eag* locus. *Science* **252**: 1560–1562.

Wheeler, D. A., C. P. Kyriacou, M. L. Greenacre, Q. Yu, J. E. Rutila, M. Rosbash, and J. C. Hall (1991). Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* **251**: 1082–1085.

Williams, J. A., and A. Sehgal (2001). Molecular components of the circadian system in *Drosophila*. *Annu. Rev. Physiol.* **63**: 729–755.

Wolf, M. E. (1999). Cocaine addiction: clues from *Drosophila* on drugs. *Curr. Biol.* **9**: R770–R772.

Wyatt, T. D. (1997). Methods in studying insect behaviour, pp. 27–56. In: *Methods in Ecological and Agricultural Entomology*. D. R. Dent, and M. P. Walton, Eds. CAB International Press, UK.

Yamamoto, D., J. M. Jallon, and A. Komatsu (1997). Genetic dissection of sexual behavior in *Drosophila melanogaster*. *Annu. Rev. Entomol.* **42**: 551–585.

Young, M. W. (1998). The molecular control of circadian behavioral rhythms and their entrainment in *Drosophila*. *Annu. Rev. Biochem.* **67**: 124–152.

Zars, T. (2000). Behavioral functions of the insect mushroom bodies. *Curr. Opin. Neurobiol.* **10**: 790–795.

Zars, T., M. Fischer, R. Schulz, and M. Heisenberg (2000). Localization of a short-term memory in *Drosophila*. *Science* **288**: 672–676.

Zerr, D. M., J. C. Hall, M. Rosbash, and K. K. Siwicki (1990). Circadian fluctuations of period protein immunoreactivity in the CNS and the visual system of *Drosophila*. *J. Neurosci.* **10**: 2749–2762.

Zhong, Y., and C. F. Wu (1991). Altered synaptic plasticity in *Drosophila* memory mutants with a defective cyclic AMP cascade. *Science* **251**: 198–201.

12

Molecular Systematics and Evolution of Arthropods

12.1. Overview

Systematics is the study of phylogeny and taxonomy. Taxonomy can be divided into descriptive taxonomy and identification. DNA is suitable for systematics studies because it provides the most direct analysis of the genetic material possible and is unlikely to be confounded by life stage or environmentally induced variability. Molecular techniques commonly used include analyses of isozymes, molecular cytogenetics, restriction analyses of DNA sequences, and DNA sequencing. Each method has virtues and limitations in the amount and type of information provided, their technical difficulties, and their costs.

There have been several significant controversies associated with using molecular techniques to study systematics and evolution. These include debates over the relative importance of molecular versus morphological data, the constancy of the molecular clock for evaluating time of divergence of taxa, the proper use of the terms *homology* and *similarity*, and the neutrality of DNA sequence variation. Another issue is how to resolve incongruencies between molecular- and morphology-based phylogenies. The immense diversity of insects and their long evolutionary history provide a challenge, but the use of molecular data provides new opportunities to discern the long and diverse evolutionary histories of arthropods and their relatives. With the use of molecular methods, systematists and population geneticists are beginning to use common approaches to study both intraspecific and interspecific genetic diversity.

12.2. Introduction

The methods and concepts used to classify arthropods, and other organisms, are themselves undergoing evolution. It all started when a formalized hierarchical system of binomial

Table 12.1. A Higher Classification of the Phylum Arthropoda

Phylum Arthropoda
Subphylum Trilobita
Subphylum Chelicerata
Class Merostomata (horseshoe crabs, eurypterids)
Class Pycnogonida (sea spiders)
Class Arachnida (spiders, mites, ticks, scorpions, phalangids)
Subphylum Crustacea (crabs, shrimp, lobsters)
Subphylum Labiata
Superclass Myriapoda
Class Diplopoda (millipedes)
Class Chilopoda (centipedes)
Class Pauropoda (pauropods)
Class Symphyla (garden centipedes)
Superclass Hexapoda
Class Parainsecta (Protura and Collembola)
Class Entognatha (Diplura)
Class Insecta (insects)

From Daly et al. (1998).

nomenclature was established by Linnaeus in 1758. In that year Carolus Linnaeus, a Swedish botanist, published *Systema Naturae* and proposed basic principles for organizing newly described species into groups and for assigning these groups to specific taxonomic categories. This resulted in a ranking classification, typically ascending from species to genus, family, order, class, superclass, subphylum, phylum, and kingdom. See Table 12.1 for a recent outline of the higher classification of the Phylum Arthropoda.

The Linnaean system made no provision for naming and classifying organisms based on evolutionary relationships, and Linnaeus assumed the living world was limited to approximately 10,000 species. The notion that a classification should be based on phylogenetic, or evolutionary, relationships developed only after Darwin's publication of *The Origin of Species* in 1859. The identification, description, and explanation of the diversity of organisms is known as systematics.

There is a broad overlap in the use of the terms **systematics** and taxonomy. Mayr and Ashlock (1991) define systematics as "the scientific study of the kinds and diversity of organisms and of any and all relationships among them" or the "science of the diversity of organisms" and taxonomy as "the theory and practice of classifying organisms." Taxonomy can be divided into descriptive taxonomy and identification (Post et al. 1992). Systematics deals with populations, species, and higher taxa. It is concerned also with variation within taxa. Thus, DNA analysis is particularly suitable for systematics studies because it is the most direct analysis of the genetic material possible and is unlikely to show life stage or environmentally induced variability.

During the 19th century, after Darwin's theory of evolution was proposed and numerous new organisms were found, ever more extensive nomenclature rules were developed to accommodate the growing numbers of plant, animal, and microbial species. So far, an estimated 1.4 million species have been identified, but these species may represent only 10% or so of the total species thought to live on this planet. Big surprises still occur, even in the relatively well-known insects. A new order of insects (called Mantophasmatodea) was discovered only recently in the mountains of Namibia in Africa (Klass et al. 2002).

Entomologists thought insect orders had all been detected until this new group was found.

Under the Linnaean system, a taxonomist begins by assessing the physical (phenotypic) characteristics that a set of species shares, then selects the most representative species to be the "type" for each genus, then the most representative genus to be the type of the family, and so on. Individual specimens are deposited in museums to serve as a reference for that species and genus. When new specimens are found with similar traits, they are categorized as part of a known species, as a new species, or as a new genus, depending on how closely the new specimens resemble the "type." This reliance on types results in dramatic changes if a systematist reevaluates a group and decides that some members don't belong. Removal of these individuals can mean that the group's name must be changed, which often is disruptive to other biologists.

In the 1980s, an assessment of methods occurred after a classification method called **cladistics**, which is based on the evolutionary histories of organisms, was proposed. The cladistics approach is based on phylogeny, while traditional Linnaean methods (phenetics) are not. Most current systematists now take the phylogenetic approach. More recently, the use of DNA-based methods has created new concerns about appropriate methods of analysis and whether molecular and traditional morphological methods provide equivalent answers. The enormous amount of DNA sequence data also requires that new methods of analysis be developed (Hall 2001).

Additional ferment in the systematics community is provided by systematists who have concluded that the fundamental Linnaean binomial system of nomenclature is obsolete (De Queiroz and Gauthier 1994, Ereshefsky 2001). Some systematists have proposed a new system called "PhyloCode" (De Queiroz and Gauthier 1994, Pennisi 2001, www.ohio.edu /phylocode). Advocates of PhyloCode want to replace the Linnaean system to make species names more stable. Under this system, genus names might be lost and species names might be shortened, hyphenated with their former genus name, or given a numeric identification. The debate over which is the better system has generated much heat, and only time will tell which approach has the fewest shortcomings (Pennisi 2001).

Systematics encompasses the study of both **phylogeny** and microevolutionary change. **Molecular evolution** encompasses: 1) analyzing the evolution of DNA and proteins and the mechanisms responsible for such changes, and 2) deciphering the evolutionary history of genes and organisms. A more recent topic, available only since the complete genomes of a variety of organisms have been sequenced, is comparative genomics.

Comparative genomics compares the overall structure and function of genomes. Molecular evolution and phylogeny are interrelated because phylogenetic knowledge is essential for determining the order of changes in the molecular characters being studied, while knowledge of the pattern and rate of change of a molecule is crucial in efforts to reconstruct the evolutionary history of a group of organisms (Li and Graur 1991, Graur and Li 2000). This chapter will introduce the most common molecular methods for these studies and describe their applications, limitations, and relative costs for systematic and evolutionary studies.

12.3. Controversies in Molecular Systematics and Evolution

Several significant controversies have been associated with using molecular tools, including: debates over the relative importance of molecular versus morphological data, the

constancy of evolutionary rates (the molecular clock), the use of the terms *homology* and *similarity*, and the neutrality of DNA sequence variation.

12.3.1. Molecular versus Morphological Traits

There is ongoing debate over whether morphological or molecular characters are *better* for constructing phylogenies (Adoutte et al. 2000). When comparisons have been made, it appears that morphological changes and molecular changes may be independent, responding to different evolutionary pressures or differently to evolutionary pressures. Hillis and Moritz (1990) noted that the real issue in choosing a technique to answer a hypothesis should be 1) whether the specific characters chosen exhibit the variation that is appropriate to the question posed, 2) whether the characters have a genetic basis, and 3) whether the data are collected and analyzed in such a way that it is possible to utilize both morphological and molecular information. Molecular and morphological data each have advantages and disadvantages.

DNA sequence data have the advantage of having a clear genetic basis, and the amount of data is limited only by the genome size (and the time and funds of the scientist!). Morphological data have the advantage that they can be obtained from fossils (if available) and preserved collections and can be interpreted in the context of ontogeny. Only limited amounts of DNA data can be obtained from preserved fossils by the PCR because of DNA degrades over time; see Chapter 8 for a discussion of PCR analyses of ancient DNA.

Moritz and Hillis (1990) conclude that the debate should not be either/or; studies that incorporate both types of data may provide better results than those using just one approach. Furthermore, some problems only can be resolved with morphological data, whereas others are better resolved with molecular data.

12.3.2. The Molecular Clock

Until the 1960s, the analysis of fossils was the only way to estimate the *time* when ancestors of extant organisms lived. Molecular studies in the 1960s provided a concept, called the **molecular clock**, that could be used to estimate the evolutionary history and time of divergence of organisms. The molecular clock was particularly useful for living species that have a poor fossil record, a very high proportion of extant species.

The molecular clock hypothesis was proposed after Zuckerkandl and Pauling (1965) examined amino acid substitutions in hemoglobin and cytochrome *c* proteins from different vertebrates. They found the rate of molecular evolution was approximately constant over time in all vertebrate lineages and concluded that amino acid sequences could be used to measure the evolutionary distance (time) between organisms by counting the number of accumulated changes (mutations).

The molecular clock is based on the assumption that basic processes such as DNA replication, transcription, protein synthesis, and metabolism are remarkably similar in all living organisms and the proteins and RNAs that carry out key "housekeeping functions" are highly conserved. Of course, over time, mutations in housekeeping genes occurred and DNA and protein sequences changed, although the changes tended to preserve the *function* of the gene rather than modify or improve it. Thus, changes in these fundamental genes should have minimal, or no, effect on function. For example, because the genetic code is

degenerate, the third base in a codon often can be altered without affecting which amino acid is designated. Changes in the code also can occur without changing protein function if amino acid changes occur in region(s) that do not affect function, or if one amino acid is replaced by a similar amino acid. The molecular clock hypothesis assumes that mutations in the housekeeping genes that constitute the clock occur at a *constant* rate, thus providing a reliable method for measuring time.

Unfortunately, analyses of different protein sequences suggest that the rates of change *can* vary between proteins and lineages, indicating that the molecular clock may tick at different rates in different lineages (Rodriguez-Trelles et al. 2001). For example, cytochrome *c* has an acceptable clockwise behavior for the original organisms studied. However, copper-zinc superoxide dismutase (SOD) behaves like an erratic clock (Ayala 1986). The average rate of amino acid substitutions in SOD per 100 residues per 100 million years is a minimum of 5.5 when fungi and animals are compared. The rate of substitutions in SOD is 9.1 amino acids/100 residues/100 million years when comparisons are made between insects and mammals, and 27.8 when mammals are compared with each other (Ayala 1986, Fitch and Ayala 1994). Thus, the molecular clock should be calibrated with data that are independently derived, and preferably with fossil evidence, if the absolute time of divergence is desired. Wilson et al. (1987) pointed out that analyses by both morphological and molecular techniques of species with abundant fossil records have reduced the uncertainty in estimating the time of divergence by several orders of magnitude. The molecular clock is thought by some to be more useful in calculating *relative* times rather than absolute times of divergence.

The molecular clock approach was used by Moran et al. (1993) to determine when endosymbiont bacteria (*Buchnera*) colonized their aphid hosts. Moran et al. (1993) compared 16S ribosomal DNA sequences of aphids and of *Buchnera* and found the clock was approximately constant. These symbiotic bacteria live within specialized aphid cells, are maternally inherited, and are essential for growth and reproduction of their hosts, indicating a long and intimate relationship. The 16S rDNA sequences indicate that the symbionts in diverse aphids are distinct and concordant with the phylogeny of their hosts, suggesting that the current distribution of *Buchnera* is due to vertical transfer from an ancestral aphid. The data also indicate that cospeciation occurred, with the aphids and their endosymbionts radiating synchronously. Moran et al. (1993) estimated the aphid and bacterial radiations occurred at a relatively constant rate, with 0.01 to 0.02 substitutions per site per 50 million years, suggesting that the association between aphids and endosymbionts began about 160 to 280 million years ago (mya).

12.3.3. The Neutral (or Nearly Neutral) Theory of Evolution

Another controversy involves the mechanism(s) of molecular evolution. At the core of the dispute is the **neutral theory of molecular evolution** (Kimura 1968, 1983, 1987, Ohta 1996, 2000b). The neutrality theory (or the modified “near neutrality” theory) recognizes that for any gene a large proportion of all possible mutations (alleles) are deleterious and that these are eliminated or maintained at a very low frequency by natural selection. The evolution of morphological, behavioral, and ecological traits is governed largely by natural selection, because it is determined by selection on favorable alleles and against deleterious ones. However, many mutations can result in alleles which are equivalent, or nearly so, to each other. These neutral mutations are *not* subject to selection because they do not affect the fitness of the individual carrying them. Neither do they affect their morphology, physiology,

or behavior. The neutrality theory states that the majority of nucleotide substitutions in the course of evolution are the result of the gradual, random fixation of neutral changes, rather than the result of positive Darwinian selection. Neutral mutations can spread in a population because only a relatively small number of gametes are sampled each generation (random genetic drift). By chance, they can be transmitted to the next generation at a higher frequency (Kimura 1968, 1983).

Ohta (1996) concluded that the "strictly neutral theory has not held up as well as the nearly neutral theory, yet remains invaluable as a null hypothesis for detecting selection." The main difference between the nearly neutral theory and the traditional selection theory is that "the nearly neutral theory predicts rapid evolution in small populations, whereas the latter predicts rapid evolution in large populations" (Ohta 1996). Kreitman (1996) noted that the neutral theory has been useful for organizing thinking about the nature of evolutionary forces acting on variation at the DNA level and has provided a set of testable predictions (acting as a useful null hypothesis). However, Kreitman (1996) argues that "the neutral theory cannot explain key features of protein evolution nor patterns of biased codon usage in certain species." Despite this, he concludes the neutral theory "is likely to remain an integral part of the quest to understand molecular evolution." Finally, both Ohta (1996) and Kreitman (1996) agree that the "nearly neutral theory" is more compatible with the current data in explaining synonymous changes and the evolution of codon bias.

Why be concerned about neutrality or nearly neutral theories? The neutrality theory is a basic assumption of some methods of estimating phylogeny, and also affects the molecular clock hypothesis (Ohta 1996, 2000b, Kreitman 1996). Data indicate that many protein, chromosome, and DNA variations *are* under selection. Data also support the hypothesis that much molecular variation is essentially (nearly) neutral. The debate thus is over how *many*, and *which*, molecular variants are selectively neutral or nearly neutral. Moritz and Hillis (1990) suggest that each molecular marker should be tested for neutrality. They also note that, because most departures from neutrality are locus-specific, selection will have relatively minor effects on analyses if many different loci are studied.

12.3.4. Homology and Similarity

A fourth issue concerns terminology. **Homology** is an important concept in biology and historically has had the precise meaning of "having a common evolutionary origin" (Reeck et al. 1987). However, homology has been used in a looser sense when comparing protein and nucleic acid sequences. Protein and nucleic acid sequences from different organisms have been called homologous when they are *similar*. According to the traditional definition of homology, amino acid or nucleotide sequences are either homologous or not. They cannot exhibit a "level of homology" or "percent homology." Reeck et al. (1987) point out that using homology to mean similarity can cause three different problems: First, sequence similarities may be called homologies, but the sequences are *not evolutionarily related*, which is certainly inconsistent. Second, similarities (again called homologies) are discussed but evolutionary origins are not, which can lead the reader to believe that coancestry is involved when it is not. Third, the similarities (called homologies) are used to support a hypothesis of evolutionary homology. The problem is that whereas similarity is easy to document, a common evolutionary origin usually is more difficult to establish, especially if fossil evidence is lacking. Several evolutionary processes other

than homology could account for sequence similarities, including **convergent evolution**, which is the independent evolution of the same characteristic in separate branches of a phylogenetic tree. When in doubt, it is better to talk about "percent similarity" of DNA sequences.

12.4. Molecular Methods for Molecular Systematics and Evolution

Systematics studies conducted prior to the 1960s primarily utilized morphological and behavioral attributes as characters, although cytogenetic characters were employed in some cases (Mayr 1970, White 1973, 1978). In the 1960s, electrophoresis of proteins began to provide new characters after Lewontin and Hubby (1966) demonstrated that protein-coding genes often are polymorphic (have more than one allele) and that gel electrophoresis of proteins could reveal the presence of functionally similar forms of enzymes (**isozymes**). Protein electrophoresis provides a relatively inexpensive method for analyzing several genes from individuals at the same time (Hames and Rickwood 1981, Pasteur et al. 1988, Murphy et al. 1996). This technique is useful for analyzing mating systems, heterozygosity, relatedness, geographic variation, hybridization, species boundaries, and phylogenetic analyses of divergences within the past 50 million years (Table 12.2).

Immunological analyses also were employed in the 1960s, but are less often used for analyses of insect systematics or evolution today. Immunological techniques provide qualitative or quantitative estimates of amino acid sequence differences between homologous proteins (Maxon and Maxon 1990).

Cytogenetic analyses of variation in chromosome structure and number have been used for studies of hybridization and species boundaries (White 1973, Table 12.2). Specific DNA sequences can be localized by *in situ* hybridization, and new staining techniques can reveal the fine structure of chromosomes by revealing banding patterns. Cytogenetic analyses are less useful for phylogenetic analyses, gene evolution, heterozygosity, and relatedness (Hillis et al. 1996).

Differences in single-copy DNA sequences revealed by **DNA–DNA hybridization** have been used for analysis of phylogenies since the late 1960s (Powell and Caccone 1990, Werman et al. 1996). DNA–DNA hybridization is employed for analyses of species and higher taxa relationships up to the family and order level but is used relatively infrequently for arthropods, so it will not be discussed further.

The expanding interest in using molecular methods for systematics and evolutionary studies is reflected by the publication of detailed protocols for molecular methods and data analysis. Hillis and Moritz (1990) provided an introduction to molecular systematics including: guidelines for sampling, collection, and storage of tissues, protocols for isozyme electrophoresis, immunological techniques, molecular cytogenetics, DNA–DNA hybridization, restriction-site analysis, nucleic acid sequencing, and analytical methods for intraspecific differentiation and phylogeny reconstruction. Pasteur et al. (1988) and Murphy et al. (1996) described protocols and methods of isozyme genetic analysis. Weir (1990) provided guidelines on analyzing population structure, phylogeny construction, and diversity using molecular and morphological data. Protocols are readily available for the PCR and nucleic acid sequencing and data analysis (Howe and Ward 1989, Doolittle 1990, Gribskov and Devereaux 1991, Hillis et al. 1990, 1996, Palumbi 1996, Green 2001, Hall 2001, Gibson and Muse 2002).

Table 12.2. Applications of Various Molecular Techniques to Systematics Problems

Research problem[a]	Isozymes	Cytogenetics	RFLP analysis	RAPD-PCR	Single-locus microsatellites	Multilocus fingerprints	DNA/RNA sequencing
Gene evolution	M	M	M	I	M	I	A
Population structure	A	M	A	M	A	I	A
Mating systems	A	M	M	M	A	I	$
Clonal detection	A	M	A	A	A	A	$
Heterozygosity	A	I	A	I or A[b]	A	M	M
Paternity testing	M	I	M	M	A	A	$
Relatedness	M	I	M	M	A	M	$
Geographic variation	A	M	A	M	A	M	A
Hybridization	A	A	A	A	M	I	$
Species boundaries	A	A	A	A	M	I	A
Phylogeny (0–5 mya)	A	M	A	I	M	I	A
(5–50 mya)	A	M	A	I	I	I	A
(50–500 mya)	M	M	M	I	I	I	A
(500–3500 mya)	I	I	I	I	I	I	A

[a] I, Inappropriate use of the technique; M; marginally appropriate or appropriate under limited circumstances; $, appropriate but probably not cost-effective; A, appropriate and effective method.

[b] May be appropriate or inappropriate if arthropod is haplo-diploid or diplo-diploid, respectively.

Modified from Hillis et al. (1996).

Protocols and results from molecular systematics are readily available: Hillis et al. (1996) updated *Molecular Systematics*; Ferraris and Palumbi (1996) edited *Molecular Zoology, Advances, Strategies and Protocols*; and Harvey et al. (1996) edited *New Uses for New Phylogenies*. Graur and Li (2000) published *Fundamentals of Molecular Evolution*, Mount (2001) published *Bioinformatics, Sequence and Genome Analysis*, Gibson and Muse (2002) provided *A Primer of Genome Science*, Hall (2001) provided a how-to manual, *Phylogenetic Trees Made Easy*, and Baxevanis and Ouellette (2001) published *Bioinformatics. A Practical Guide to the Analysis of Genes and Proteins.*

12.4.1. Protein Electrophoresis

The term **isozyme** is a general designation for multiple forms of a single enzyme. Isozymes will catalyze the same reaction, but may differ in properties such as the pH or substrate concentration at which they best function. Isozymes are complex proteins made up of paired polypeptide subunits; their subunits may be coded for by different loci. For example, protein Z could be a tetramer made up of two polypeptides, A and B. Five isozymes of protein Z could exist and be symbolized AAAA, AAAB, AABB, ABBB, and BBBB. Isozymes may have different isoelectric points and be separated by gel electrophoresis.

The term **allozyme** refers to variant proteins produced by *allelic* forms of the *same locus*. Thus, A is now A'. A different mutation of A could produce A". Allozymes are a *subset* of isozymes; allozymes may differ by net charge or size so they can be separated by electrophoresis.

The process of analyzing isozymes or allozymes can be divided into five steps: extraction of proteins, separation, staining, interpretation, and application. Proteins are more difficult to handle than DNA because they are more susceptible to degradation. Proteins must be frozen and stored at −70°C, but even at those temperatures some proteins can degrade within months.

Proteins are separated in an electric field on a gel. In gels with a single pH, the proteins move through the gel at a continuous rate, but in gels with a pH gradient, they move until they reach their isoelectric point and then stop. The resultant electrophoretic bands are visualized by appropriate staining (Murphy et al. 1996, May 1992). If a general protein detection system is used, only those proteins present in large quantities are detected, but more specific stains can be used. Specific stains and buffer recipes are available for more than 50 enzymes (May 1992). The banding phenotypes observed on the gels can be interpreted in terms of genes and their alleles (Pasteur et al. 1988, May 1992).

Protein-coding genes are often codominant, with both alleles being expressed in heterozygous organisms. This makes it possible to relate a particular phenotype to a given genotype, if we assume that isozyme data reflect changes in the DNA sequence. To interpret the banding patterns, the number of subunits in the enzyme and the distribution of enzymes in particular cells or tissues should be known (May 1992).

Analyses of isozymes remain cost-efficient and useful for deciphering the systematics, population genetics, and evolution of insects (Table 12.2). Protein electrophoresis can be conducted using starch (horizontal or vertical gel systems), polyacrylamide, agarose, and cellulose acetate gels as substrates (Hames and Rickwood 1981). Each has specific advantages and disadvantages (Moritz and Hillis 1990). However, isozyme or allozyme data are useful for estimating the evolution of only a portion of the genome: those genes coding for enzymes that have a different charge and size. The data also are most useful

for analyzing relatively closely related taxa. Unfortunately, allozyme variation in some insects, such as aphids and Hymenoptera, is low, and other molecular techniques are required.

An example of allozyme analysis illustrates an economically important application of the technique. Twenty-four populations consisting of three subspecies of *Culicoides variipennis* (Diptera, Ceratopogonidae) from different geographic regions were examined for genetic differences (Tabachnick 1992). Twenty-one loci were examined among the 24 populations of this vector of bluetongue virus, a disease that causes losses of $125 million annually to the U.S. livestock industry. The results were analyzed with a stepwise discriminant analysis and are consistent with the conclusion that there are three North American subspecies; the three subspecies may even be sufficiently differentiated to be considered species. Furthermore, geographic variation in bluetongue disease epidemiology is correlated with the distributions of the three subspecies. These, and subsequent, data support the hypothesis that one subspecies is more effective as a vector of the virus (Tabachnick 1996, Holbrook et al. 2000). The results could have implications for pest management programs and significant economic impacts. Thus, the areas inhabited only by the two (nonvector) subspecies could be considered virus-free regions, and animals raised in such areas would not have to undergo extensive testing when livestock or germ plasm from them is exported to regions without the disease.

12.4.2. Molecular Cytology

Three breakthroughs in cytogenetic techniques revived this approach to systematic and evolutionary studies. The first was the discovery that hypotonic treatment spreads metaphase chromosomes, allowing more accurate counts of chromosome numbers and details of chromosome morphology. The second was the development of chromosome banding techniques that allow the identification of specific types of DNA within homologous chromosomes. The third was the development of ***in situ* hybridization** techniques, which allow specific DNA sequences to be localized to particular segments of the chromosomes.

In situ hybridization involves annealing single-stranded probe molecules and target DNA to form DNA duplexes. *In situ* hybridization is effective in locating satellite DNA, ribosomal gene clusters, or duplicated genes of polytene chromosomes and can even locate single-copy DNA on mitotic chromosomes. Chromosomal DNA is denatured in such a way that it will anneal with high efficiency to complementary ss nucleic acid probes to form hybrid duplexes. Because chromosomal DNA is complexed with proteins and RNA, the efficiency of *in situ* hybridization is determined by how well the chromosomal DNA can be denatured, how much DNA is lost during fixation and treatment, and whether chromosomal proteins are present in the region of interest (Sessions 1996). Sites where hybridization between a radioactive probe and its target DNA occur are visualized by autoradiography or by using nonradiographic labeling techniques such as biotinylation.

Chromosome morphology may be used as a taxonomic character (Table 12.2). In many cases, chromosomes can be identified by their relative size, centromere position, and secondary constrictions. Many chromosomes, particularly insect polytene chromosomes, have complex patterns of bands or other markers that can be used to identify specific populations or to discriminate between closely related species. Distinctive patterns can be obtained by Q-, G-, or C-banding that identify chromosomes in most species.

Q-banding is the simplest technique and involves treating chromosome preparations with quinacrine mustard or quinacrine dihydrochloride, which produces fluorescent bands

that are brightest in AT-rich regions of the chromosomes. Q-banding is visible only with UV optics, and the bands fade rapidly. **G-banding** involves treating chromosome preparations with trypsin or NaOH and staining with Giemsa in a phosphate buffer, which yields alternating light and dark bands. The dark bands are primarily AT-rich regions and thus correspond to most Q-bands. **C-banding** requires a stringent extraction step that can result in loss of chromosomal DNA. During C-banding, chromosomes are treated with a strong base at a high temperature, incubated in a sodium citrate solution again at high temperature, and stained in a concentrated Giemsa solution. C-banding extracts almost all of the non-C-band chromatin, leaving only constitutive heterochromatin, which usually contains rapidly reassociating repeated DNA sequences (Sessions 1996).

Cytogenetic data provide information independent from morphological, biochemical, or behavioral data for phylogenetic analyses. Cytogenetic data can reveal differences or similarities that may not be obvious at the morphological level. Chromosome size, shape, number, and ploidy levels can provide insights into the genetic architecture of taxa. Banding studies reveal aspects of the structural organization of chromatin on individual chromosomes, whereas probes of DNA sequences with *in situ* hybridization can reveal finer details of chromosome anatomy in terms of spatial arrangement, as well as the presence or absence of particular kinds of DNA sequences.

12.4.3. Restriction Fragment Length Polymorphism (RFLP) Analysis

Restriction enzyme analyses are versatile, providing information on the nature, as well as the extent, of differences between sequences in nuclear or mitochondrial DNA (Dowling et al. 1996, Table 12.2). RFLP analysis reveal variations within a species in the length of DNA fragments generated by a specific restriction endonuclease. RFLP variations are caused by mutations that create or eliminate recognition sites for the restriction enzymes.

RFLP analyses can be used effectively, and relatively economically, to analyze clonal populations, heterozygosity, relatedness, geographic variation, hybridization, species boundaries, and phylogenies ranging in age from 0 to 50 million years ago (mya) (Table 12.2). It is possible to analyze more loci per individual by RFLP analysis than by DNA sequencing because RFLPs are less time consuming and expensive. The information provided for each locus is less complete (Dowling et al. 1996, Hall 1998). Higher-level systematics studies only rarely have used RFLPs.

More than 1400 restriction enzymes are known that cut DNA at a specific position within a specific recognition sequence. See Chapter 5 for a discussion of restriction digests, as well as Brown (1991), and catalogues from a variety of commercial producers. Such recognition sequences are typically 4 to 6 bp long, although they can be as large as 12 bp. The specificity of restriction enzymes means that a complete digestion will yield a reproducible array of DNA fragments. Changes in the number and size of fragments can occur by changes in DNA sequence by rearrangements (inversions, tandem duplication, inverted duplication), or addition, deletion, or substitution of specific bases.

Once the DNA is digested with a restriction enzyme, the fragments produced are sorted by size using agarose or polyacrylamide gel electrophoresis. DNA fragments of known length are run on each gel to serve as an internal standard and to allow the size of the experimental fragments to be estimated. The DNA fragments in the gel are visualized by several methods, including staining with ethidium bromide (if the DNA was previously amplified by the PCR)

or probing Southern blots with labeled probes. The detection technique employed depends on the amount of DNA present in the gel.

Staining with ethidium bromide is simple and cheap, but least sensitive. The minimal amount of DNA in a band that can be detected by ethidium bromide is about 2 ng, so small fragments can be detected only if a large amount of DNA is present. DNA probes can be end labeled by adding ^{32}P-labeled nucleotides to the ends of DNA probes produced by the restriction enzymes. Intensity of labeling is independent of fragment size and is more sensitive than EtBr, with 1 to 5 ng of DNA easily visualized. If primers are available, DNA can be first amplified by the PCR, cut with a restriction enzyme, and labeled by ethidium bromide.

If less DNA is available, radiolabeled DNA probes can be used to visualize fragments. Southern blot hybridizations are highly sensitive, and picogram quantities of DNA can be detected, although small fragments less than 50 bp are more difficult to detect. Southern blots require a suitable probe with sufficient sequence similarity to the target DNA that a stable hybrid can be formed at moderate to high stringency. The use of probes from other species (heterologous probes) makes interpretation of results more difficult.

12.4.4. DNA Sequencing

Sequences of proteins, RNA, and DNA have been obtained only during the past 40 years. The first sequence information was obtained from proteins in the mid-1950s. RNA was sequenced in the mid-1960s, and DNA sequences were obtained in 1975 after DNA sequencing methods were developed. Techniques for DNA sequencing were described in Chapters 7 and 8 and are available from many sources (Innis et al. 1990, Hillis et al. 1996). The use of the PCR reaction makes DNA sequencing less time consuming and expensive for systematic studies, and the availability of core facilities and commercial resources that can conduct automated sequencing has reduced the need for individual laboratories to carry out their own sequencing reactions.

DNA sequence data can be used to: 1) construct molecular phylogenies to evaluate the evolution of particular genes or gene families, 2) evaluate evolutionary changes within species, and 3) construct phylogenies of different species. DNA sequences can be obtained for single-copy genes, mitochondrial (mt) DNA, and ribosomal DNA. Sequences can be used to study most systematics problems from intraspecific variability to phylogeny of all organisms (Table 12.2). Sequence data is appropriate for analysis of intraspecific variation, cryptic species, geographic variation, reproductive behavior, and heterozygosity estimates. However, DNA sequencing remains relatively expensive and time-consuming and may have limited use if very large numbers of individuals must be analyzed. Sequence analysis of nuclear or mt DNA sequences provides very large amounts of detailed data. The number of potential characters that can be examined theoretically is limited only by the number of nucleotides in the DNA of the organism. Declining costs of DNA sequencing could make sequencing more commonly used in the future.

12.4.5. Fragment Analyses of Genomic DNA

Fragment analyses, which include RAPD-PCR, single-locus microsatellites or multilocus DNA fingerprinting, can be used for some systematics problems (Table 12.2). The random amplified polymorphic DNA method of the PCR (described in Chapter 8) has been used to discriminate between cryptic sympatric species. Multiple RAPD markers may have to be

employed to produce a banding pattern that can be analyzed by discriminant analysis, although the need to conduct multiple RAPD reactions would make RAPD-PCR more expensive and time consuming. RAPD-PCR may be useful for examining hybridization and species boundaries, as well as clonal variation.

Single-locus microsatellites are potentially useful for analysis of population structure, mating systems, clonal boundaries, heterozygosity, paternity testing, relatedness, and geographic variation (Table 12.2). Multilocus DNA fingerprinting can be used for clonal detection and paternity testing.

12.5. Targets of DNA Analysis

Sequence analyses of nuclear, mitochondrial, and ribosomal DNA have been employed in systematics studies, as have microsatellites and introns (Caterino et al. 2000). The following discussion describes some of the attributes of these targets, which are relevant because specific assumptions may be essential for an appropriate phylogenetic analysis method.

12.5.1. Mitochondria

Mitochondria are the cell's respiratory power plant for the generation of ATP. Mitochondria are thought to have developed more than a billion years ago when a free-living eubacterium took up residence within another cell (Margulis 1970). Sequence analysis of modern mitochondrial DNA suggests that α-Proteobacteria, such as *Rickettsia*, *Anaplasma*, and *Ehrlichia*, are the closest contemporary relatives of that eubacterium (Gray et al. 1999, Lang et al. 1999). The relatively low gene content of mtDNA, compared with even the smallest eubacterial genome, suggests that loss or transfer of genetic information occurred at an early stage in the evolution of the "protomitochondrial genome" (Gray et al. 1999). What is not clear is whether mt originated as a result of a single endosymbiotic event or more than one (Lang et al. 1999).

Mutation rates in mtDNA are variable within the eukaryotes, with mammal mtDNA having a mutation rate at least 50 times greater than mitochondria in plants (Lang et al. 1999). The number and type of genes present differ in the eukaryotes, with perhaps seven to 10 independent losses having occurred for each gene. The evolution of mitochondrial genes over the past billion years has been complex; some genes apparently were transferred independently several times into the nuclear genome; some genes were lost without transfer to the nuclear genome because of gene substitution; genes were acquired by lateral gene transfer, as well (Gray et al. 1999). However, there is no evidence that mt were transferred between different eukaryotes.

There is no evidence that once genes transferred to the nucleus they were regained by the mitochondria, and there is no evidence for widespread and substantial lateral transfer of genetic information into or between mitochondria (Lang et al. 1999). The predominantly vertical inheritance of genes from mitochondrion to mitochondrion is a prerequisite for phylogenetic analyses. Within mitochondria, there are regions that diverge rapidly, while other regions are highly conserved, making the different regions suitable for analysis of different taxonomic levels (Simon et al. 1991, Liu and Beckenbach 1992, Tamura 1992, Caterino et al. 2000).

Animal mitochondria are small (16 to 20 kb in length), circular, and lack introns, with the genes compactly arranged on both DNA strands. With a few exceptions, animal mtDNA

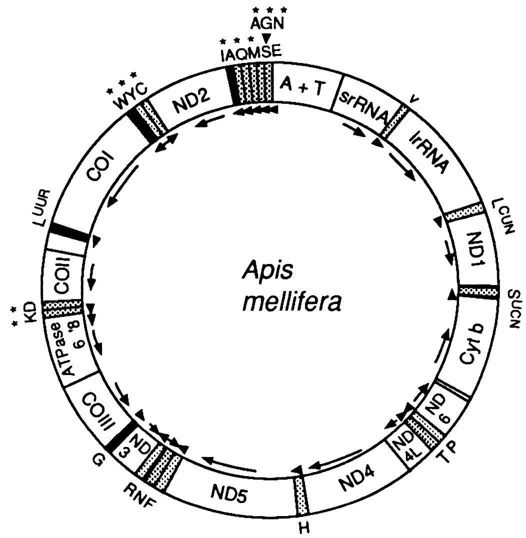


Figure 12.1. Map of mitochondrial DNA of the honey bee *Apis mellifera*. Genes for tRNAs are denoted by the one-letter code for their corresponding amino acids. tRNA genes with asterisks are in a different position compared with the same genes in the genome of *Drosophila yakuba*. Protein-coding genes are denoted COI, COII, COIII for the genes encoding subunits 1, 2, or 3 of *cytochrome C oxidase*, *Cyt b* for the *cytochrome b* gene, and ND4L for the genes encoding subunits 1-6 and 4L of the NADH dehydrogenase system. The AT-rich region containing the origin of replication is denoted A+T. The direction of transcription for each coding region is shown by arrows. (Redrawn from Crozier and Crozier 1993.)

contains the same 37 genes coding for small- and large-subunit rRNAs, 13 proteins, and 22 tRNAs arrayed in an order that is well conserved within phyla (Figure 12.1). All mtDNAs have at least one noncoding region, which contains regulatory elements for replication and transcription, but intergenic sequences are small or absent (Boore and Brown 1998). The control region containing the origin of replication is extremely rich in adenine and thymine in insects. Within the insects, the tRNA genes are known to vary in position between the orders Diptera and Hymenoptera, and within the Diptera (Crozier and Crozier 1993). There are thousands of mitochondria in each cell, so mtDNA is abundant and relatively easy to obtain, even from somewhat degraded samples. By contrast, nuclear genes evolve more slowly, which makes it possible to extend the analysis further into the past.

Mitochondrial DNA can be used in evolutionary studies, including analyses of population structure and gene flow, hybridization, biogeography, and phylogenetic relationships (Avise et al. 1987, Lang et al. 1999). The small size, relatively rapid rate of evolutionary change, and (usually) maternal inheritance of mtDNA make it suitable for examining population history and evolution among closely related taxa (Gray 1989, Lansman et al. 1981, Simon et al. 1991, Caterino et al. 2000), as well as deeper evolutionary relationships (Lang et al. 1999, Caterino et al. 2000). Molecular studies of mtDNA have employed study of RFLPs and sequencing of specific regions of the mtDNA following cloning or amplification by the PCR (Satta and Takahata 1990, Pashley and Ke 1992, White and Densmore 1992). Gene order also can be used as a phylogenetic tool (Boore and Brown 1998).

The lack of recombination in mitochondria means that fixation of an advantageous mutation by selection will fix all other polymorphisms by "genetic hitchhiking." Even the quickly

evolving noncoding origin of replication region may not have neutral allele frequencies because this region is linked to the rest of the genome (Ballard and Kreitman 1995). Ballard and Kreitman (1995) point out that violation of the neutral evolution assumption can have important phylogenetic implications. It violates a major assumption of one method (UPGMA clustering, described later). Furthermore, selection or parasite-induced sweeps can mimic the effects of interpopulation migration or population bottlenecks. Statistical methods can detect which DNA sequences are under positive selection (Mooers and Holmes 2000, Yang and Bielawski 2000).

Mitochondria have been used as "molecular clocks" to time the divergences of organisms from each other. However, dating of evolutionary events can be problematic when nonneutral evolution within species is combined with altered rates of evolution in the sister taxon. Thus, Ballard and Kreitman (1995) suggest that phylogenetic and statistical tests of neutrality be carried out when using mtDNA in phylogenetic analysis. Furthermore, in many cases, it appears that mitochondrial molecular clocks tick at different rates in different lineages and at different times within a lineage.

There are difficulties in working with mtDNA. These include the lack of recombination, which makes mtDNA essentially a single heritable unit. Although recombination in mtDNA has not been observed in insects, it has been found in fungi when **heteroplasmy** (presence of two types of mitochondria) occurs (Saville et al. 1998). Lack of recombination potentially produces gene diversity estimates that have larger standard errors than those determined using nuclear loci that can recombine. Biparental inheritance occurs occasionally in insects, which can complicate population studies (Lansman et al. 1983, Kondo et al. 1990, Matsuura et al. 1991). Introgression of mitochondria between *Drosophila* species has been suggested as an explanation for the presence of mitochondria from a related species (Aubert and Solignac 1990).

The complete mitochondrial sequences of a number of insects and other arthropods are available in GenBank and other databases (Caterino et al. 2000, Table 12.3). A map of *Drosophila yakuba* mitochondria is shown in Chapter 3 (Figure 3.6), and a map of the mtDNA of the honey bee is shown in Figure 12.1. Honey-bee mtDNA is 16,343 bp long, and 11 of the tRNA genes are in altered positions compared to their positions in *D. yakuba* (Crozier and Crozier 1993).

12.5.2. Ribosomal RNA

Ribosomes are a major component of cells that are involved in translating messenger RNA into proteins. Ribosomes consist of ribosomal RNA (rRNA) plus proteins. All ribosomes can be dissociated into two subunits, a large and a small, each containing rRNA and protein molecules. The larger subunit (28S) may contain a smaller RNA molecule in addition to a larger RNA. Ribosomal RNAs are used frequently to evaluate evolutionary relationships among species because they are universally present. Ribosomal RNAs contain regions that are conserved and regions that are more variable, so rRNAs can serve as both slow and fast clocks.

In eukaryotes, the genes encoding the 18S (small subunit) and 28S (large subunit) rRNAs are clustered as tandem repeats in the nucleolus-organizing regions of the nuclear chromosomes (Figure 12.2), but two ribosomal genes are found in mitochondria (Figure 12.1). In most animals, there are 100 to 500 copies of rDNA in the nuclear genome in tandemly repeated transcription units. Ribosomal gene copy number ranges from as few as 45 in the

Table 12.3. Complete Sequences of Mitochondria from the Arthropoda and Location of Mitochondrial Databases

Species	Class, Order	References[a]
Anopheles gambiae	Insecta, Diptera	Beard et al. 1993
An. quadrimaculatus	Insecta, Diptera	Mitchell et al. 1993
An. sinensis	Insecta, Diptera	Hwang et al. 2001
Apis mellifera	Insecta, Hymenoptera	Crozier and Crozier 1993
Armadillium vulgare	Crustacea, Isopoda	Hwang et al. 2001
Bombyx mori	Insecta, Lepidoptera	Lee et al. (GenBank NC 002355)
Ceratitis capitata	Insecta, Diptera	Spanos et al. 2000
Cochliomyia hominivorax	Insecta, Diptera	Lessinger et al. 2000
Daphnia pulex	Crustacea, Cladocera	Crease 1999
Drosophila melanogaster	Insecta, Diptera	Lewis et al. 1995
Drosophila yakuba	Insecta, Diptera	Clary and Wolstenholme 1985a,b
D. melanogaster, simulans mauritiana, and sechellia	Insecta, Diptera	Ballard 2000
Heterodoxus macropus	Insecta, Pthiraptera	Shao et al. 2001
Limulus polyphemus	Chelicerata	Lavrov et al. 2000
Lithobius forficatus	Myriapoda	Hwang et al. 2001
Locusta migratoria	Insecta, Orthoptera	Flook et al. 1995
Macrobrachium nipponense	Crustacea	Hwang et al. 2001
Megaphylum sp.	Myriapoda	Hwang et al. 2001
Tetrodontophora bielanensis	Insecta, Collembola	Nardi et al. 2001
Triatoma dimidiata	Insecta, Hemiptera	Dotson and Beard 2001
Mitochondrial databases;	MitBASE	Attimonelli et al. 2000
	AMmtDB	Lanave et al. 2002
	Variable number of tandem repeats in mt DNA	Lunt et al. 1998
	Lepidopteran control region	Taylor et al. 1993
	PCR of long sections in 14 orders	Roehrdanz and DeGrugillier 1998
	12S rRNA	Hickson et al. 1996
	PCR primers	Simon 1991, Simon et al. 1990
	Alignments of 13 orders	Buckley et al. 2000

[a] See also GenBank for sequences at: www.ncbi.nim.nih.gov:80/entrez

fly *Sciara coprophila* to more than 3000 in the grasshopper *Locusta migratoria*. In a survey of 30 species of mosquitoes, copy numbers ranged from 39 to 1023 (Kumar and Rai 1990).

The repeated transcription unit is composed of a leader promoter region known as the External Transcribed Spacer (ETS), an 18S rDNA coding region, an Internal non-coding Transcribed Spacer region (ITS), a 28S rRNA coding region, and an InterGenic nontranscribed Spacer segment (IGS) (Figure 12.2). Different portions of the repeated transcription unit evolve at different rates in the nuclear genome. Thus, evolutionary studies employ analysis of different segments, depending on the taxonomic level being studied. In general, a higher degree of polymorphism has been found in the *noncoding segments* (ETS, ITS, IGS). The most variable part of the repeated unit is the intergenic spacer (IGS), which typically contains reiterated subrepeats ranging from about 50 to several hundred base pairs in length (Cross and Dover 1987). The coding regions of the repeated unit change

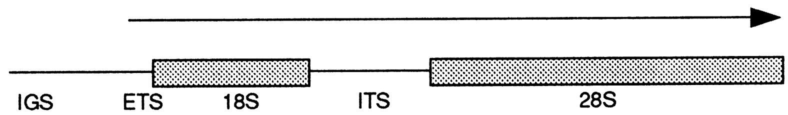


Figure 12.2. A simplified diagram of the ribosomal DNA repeat unit of eukaryotes. IGS is the intergenic spacer region. ETS is the external transcribed spacer, and 28S is the large-subunit rRNA gene. The arrow indicates the direction of transcription. Most insects have hundreds of ribosomal RNA genes in tandem array. Some will contain R1 and R2 retrotransposable elements (not shown) in specific locations. Genes with R1 and R2 elements will produce nonfunctional message.

relatively little and can be used for systematic studies of higher level taxa or for ancient lineages. Highly conserved regions are no doubt important for maintaining the characteristic secondary and tertiary structure of rRNA molecules (Simon et al. 1991, Van de Peer et al. 1993, Caterino et al. 2000).

Ribosomal RNA genes undergo concerted evolution so that the sequence similarity of members of an RNA family is expected to be greater within a species than between species. Unequal crossing over, gene conversion, and illegitimate recombination are responsible for concerted evolution. Ribosomal gene families are considered to be "quite uniform" (Ohta 2000a).

Two retrotransposons, called R1 and R2, have been found in the 28S rRNA genes of most insects (Eickbush 2002). These elements are thought to have been associated with arthropods for more than 500 million years, and usually they are precisely located at the same nucleotide position within the 28S rRNA gene. Most R2 elements are located about 74 bp upstream from the site of R1 insertions. R1 and R2 elements lack long terminal repeats (LTRs) and block the production of functional rRNA (Eickbush 2002). The insect host survives because it contains hundreds of rRNA genes and the R2 elements are kept from invading too many of them, through unknown mechanisms. Surprisingly, most R1 and R2 elements have not accumulated mutations that would make them inactive. Some species have more than one family of R1 or R2 elements, and sequence identity between the different families can be low, suggesting either that each insertion family is able to maintain its copy number without eliminating other families, or that there has been horizontal transfer of R1 and R2 elements between species. A phylogenetic analysis of R2 elements suggests that multiple lineages of R2 elements have evolved in arthropods and these have been differentially maintained (Eickbush 2002).

The relative stability of R2 elements is thought to be due to their presence in a specific site within the rRNA genes. It is likely that R2 elements are lost but new copies are produced by retroposition. It is unknown whether the elements are maintained because R2 elements can monitor their copy number and expand when their copy number declines, or whether they insert at continuously high rates but are restrained by selection on their host.

12.5.3. Satellite DNA

Satellite DNA may consist of a large fraction of the total DNA in an insect. Microsatellites are usually species specific, perhaps because this DNA evolves at a very high rate. There are only a few cases in which the same satellite sequences have been found throughout an entire genus. Satellite DNA can be used for species diagnoses or analyses of populations

(Bachmann et al. 1993, Caterino et al. 2000). Satellite DNA has most often been used in population ecology and will be discussed in Chapter 13.

12.5.4. Introns

Introns within single-copy nuclear genes may be used in systematics studies. Such noncoding regions are perceived to be highly variable (Caterino et al. 2000). The use of introns to resolve origins of invasive populations of the Mediterranean fruit fly is described in Chapter 13.

12.5.5. Nuclear Protein-Coding Genes

A variety of protein-coding loci have been used in molecular systematics of insects (Friedlander et al. 1992, Caterino et al. 2000). These include *α-amylase*, *acetylcholine esterase, actin*, *alcohol dehydrogenase*, *arylphorin*, *cecropin*, *chorion* genes, *dopa decarboxylase*, *elongation factor-1-alpha*, *esterase*, *glycerol-3-phosphate dehydrogenase*, *glycerol-6-phosphate dehydrogenase*, *guanylate cyclase*, globin family genes, histones 1 and 4, *hunchback*, *kruppel*, *luciferase*, lysozyme intron, myosin alkali light chain intron, *nullo*, *opsin*, *period*, *phosphoglucose isomerase*, *phosphoenolpyruvate carboxykinase*, *prune*, *resistance to dieldrin, Cu*, *Zn-superoxide dismutase*, sodium channel para locus 1, *snail*, *timeless*, *triosephosphate isomerase*, *vestigial*, *white*, *wingless*, *xanthine dehydrogenase*, yolk protein 1 and 2, and *zeste* (Caterino et al. 2000). Nuclear genes exhibit a wide range of evolutionary rates.

Problems with nuclear DNA sequences used for phylogenetic analysis include the fact that they may be heterozygous; they also are present in low copy number, which may make them difficult to amplify by the PCR. Furthermore, many genes contain large introns that make it difficult to amplify more than one exon unless reverse transcriptase PCR (RT-PCR) is carried out on mRNA. Caution is also warranted: many single-copy loci actually are present in more than one copy. Furthermore, pseudogenes (inactive forms of a gene) may create problems if comparisons are made inadvertently between genes and pseudogenes.

12.5.6. Rare Genomic Changes

DNA sequence data are used most often to construct phylogenies. However, Rokas and Holland (2000) suggest that single nucleotide substitutions may not always be informative and argue that rare genomic changes such as intron **indels** (an insertion or deletion), retroposon integrations, signature sequences, mitochondrial and chloroplast gene order changes, or gene duplications and genetic code changes provide useful information with "enormous potential for molecular systematics."

As an example, Rokas and Holland (2000) review research conducted to resolve the relationship of the Strepsiptera, Diptera, and Coleoptera. Strepsipteran fore wings resemble the hind wing balancing organs (halteres) of Diptera. Under one scenario, dipteran (hind wing) halteres could be homologous to the fore wings of Strepsiptera if a homeotic mutation reversed the position of the structures in Strepsiptera. By contrast, some would place the Strepsiptera closer to the Coleoptera because both use the hind wings for flight. Analysis of 18S rDNA sequence data did not resolve the question. However, a unique intron insertion was found in the homeobox of the *engrailed* gene of Diptera and Lepidoptera, which is absent from other insects and other outgroups. If Strepsiptera had the intron, it would

support a sister group relationship with Diptera, but its absence would not. Cloning of the strepsipteran *engrailed* gene revealed the intron was absent, indicating that halteres of Strepsiptera and Diptera are more likely a case of convergent evolution.

Another example in which a rare genomic change may provide useful phylogenetic information involves the gene order in mitochondria of insects, crustaceans, and myriapods (Boore et al. 1998). The mitochondria of both crustaceans and insects share a changed gene order, suggesting that myriapods are an outgroup.

12.6. Steps in Phylogenetic Analysis of DNA Sequence Data

An ongoing need in molecular systematics and evolution studies is to resolve which genes are informative for which questions. The wealth of information obtained from DNA sequences can provide insights into evolution and speciation, but how does one choose appropriate genes for a specific problem? How can estimates of genetic distance be used to make judgments about species status or date of speciation events? Several concepts are important in using various phylogenetic programs to resolve the relationships of different taxa (Gibson and Muse 2002).

12.6.1. Gene Trees or Species Trees

Phylogenetic analysis of a particular locus may not agree with the species phylogeny (Caterino et al. 2000). This may be due to the horizontal movement of genes, to duplication and extinction of one of the genes, or to lineage sorting (deep coalescence). Mitochondrial genes may be more reliable than nuclear genes for evaluating some recent divergences (Caterino et al. 2000). Methods for inferring species trees from multiple gene trees have been developed and can be carried out with a program called GeneTree (Caterino et al. 2000).

12.6.2. Rooted or Unrooted Trees

Most phylogenetic methods can produce only unrooted trees. Information regarding evolutionary rates or the most ancient relationships is needed to root the inferred trees. A comparison of the two concepts is shown in Figure 12.3. For any four taxa (the tips of the branches) there are three distinct unrooted trees (I, II, III). Each unrooted tree can be rooted on any of its five branches; two of the possible rooted trees for the center unrooted tree are shown (IIA, IIB).

12.6.3. Tree Types

The immense diversity of insects and their long evolutionary history provide a challenge for systematists. Because of mutation, high reproductive rates, natural selection, and stochastic events, populations change through time. A process of gradual change through thousands of years can result in a different species; a change within a single lineage is called **phyletic speciation**. Speciation also can occur through **cladogenic speciation**, in which two populations of a species become isolated and diverge genetically as a result of independent mutation, natural selection, and genetic drift. Other models for speciation include speciation through hybridization and polyploidy, or by modification of regulatory genes.

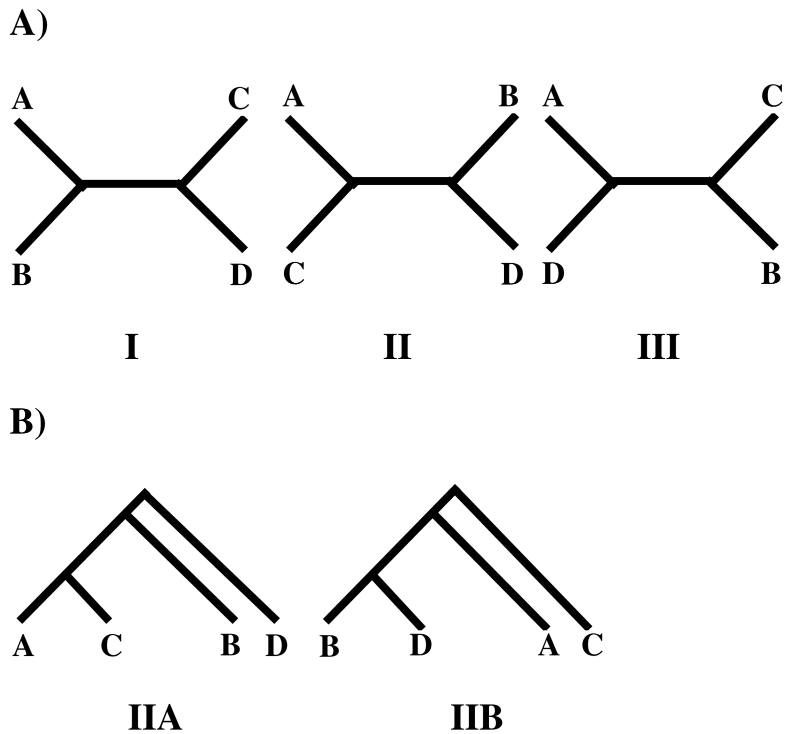


Figure 12.3. For any four taxa (A, B, C, D; the tips of the branches), there are three different unrooted trees (I, II, and III at top of figure). Each unrooted tree can be rooted on any of its five branches; two of the possible five rooted trees for unrooted tree II are shown at the bottom (IIA, IIB). Rooted trees II C, IID, and IIE are not shown. (Redrawn from Gibson and Muse 2002.)

The attributes of an organism used by systematists to establish their relation to other organisms are called **characters**. Characters can be based on morphology, physiology, ecology, behavior, biochemistry, or genetics.

There are several approaches to developing classifications using these characters. Unfortunately, the debates over which approach is more objective, appropriate, or practical have not been resolved. **Phenetic systematics** focuses on overall similarities among organisms, involves all possible characters, and calculates average similarities with all characters assumed to be equally useful. In some cases, classifications based on phenetic similarities may reflect the phylogeny of taxa because those that are most similar may well have shared a most recent ancestor, but this need not be so because of convergent evolution.

Cladistic (phylogenetic) systematics uses only cladistic relationships as a basis for constructing classifications (Hennig 1966). The rate or amount of change is not considered, and only monophyletic taxa are allowed. This approach focuses on the order of origin of lineages and also takes into account the amount and nature of evolutionary change which occurs after cladogenesis. Characters are not assumed to be equal and are weighted accordingly. One of the major difficulties in any reconstruction of phylogeny is to determine which character is primitive or ancestral (**plesiomorphic**) and which is derived (**apomorphic**).

Classifications often are represented in graphical forms as treelike dichotomous branching graphs or **dendrograms**. A dendrogram produced from phenetic information is called a **phenogram**. A phenogram shows how similar the group is, but does not provide information about probable lines of descent. When a dendrogram is produced from cladistic information it is called a **cladogram**. A cladogram shows the sequence of origin of clades and indicates the times at which the various cladogenic events have taken place. If the dendrogram includes both phenetic and phylogenetic data, it is a phylogram, or **phylogenetic tree**, and indicates not only the cladistic branching but the relative amount of change that has occurred. Those species which show the closest relationship are grouped together into larger, more inclusive groups or genera. Genera are grouped into families, and families into orders, classes, and phyla.

12.6.4. Project Goals and Appropriate DNA Sequences

The first step is to consider carefully the project goals and to evaluate published information as to which genes/DNA sequences may be most appropriate to answer the question (Figure 12.4). Analysis of the evolution of orders will require different approaches than analysis of species within a genus. Whether a specific DNA sequence is appropriate for a particular project is difficult to predict in advance unless a survey has been, or can be, conducted to determine if the appropriate level of variability is present.

Analyses of different DNA sequences provide information about different levels of phylogenetic analyses over a broad range of taxa. For example, ribosomal genes are widely used because they are highly conserved but have regions that change rapidly and regions that change slowly. Once the target gene or other DNA sequence has been chosen, primers must be designed or made to amplify the target sequence by the PCR or a cloning strategy developed. Cloning of target DNA may be required if inadequate sequence information is available in the literature or GenBank and if "universal primers" are unavailable.

Once insects have been obtained by collecting or from museums, DNA must be extracted. As noted in Chapter 8, PCR results vary with the preservation method. Dried museum specimens are likely to contain degraded DNA, so using DNA sequences that are present in multiple copies (such as mitochondrial or ribosomal DNA) may be more appropriate than using single-copy nuclear genes. DNA extraction results are better with freshly collected, frozen (at −80°C), or alcohol-preserved (95% EtOH) insects.

The sequencing outcome depends upon the purity of the DNA employed and the fidelity of the sequencing procedures. It is important to sequence both strands to avoid errors. Once sequence information has been obtained, analysis can provide several types of information, including possible structure, function, and characteristics of the protein. The similarity of the sequences to sequences obtained from other organisms can be compared. The tasks of collating, assembling, and correcting the sequence data are usually performed with the help of a variety of computer programs (Gribskov and Devereux 1991, Fortna and Gardiner 2001).

Analysis of sequence data involves several steps (Figure 12.4). Figure 12.5A shows the form in which sequence data is obtained from a sequencing laboratory. Figure 12.5B shows the data after they have been analyzed to show the open reading frame (ORF) of the coding strand (the line with the •), the codons, and the region at the end of the sequence that represents the vector DNA (which should be excluded from the subsequent analysis).

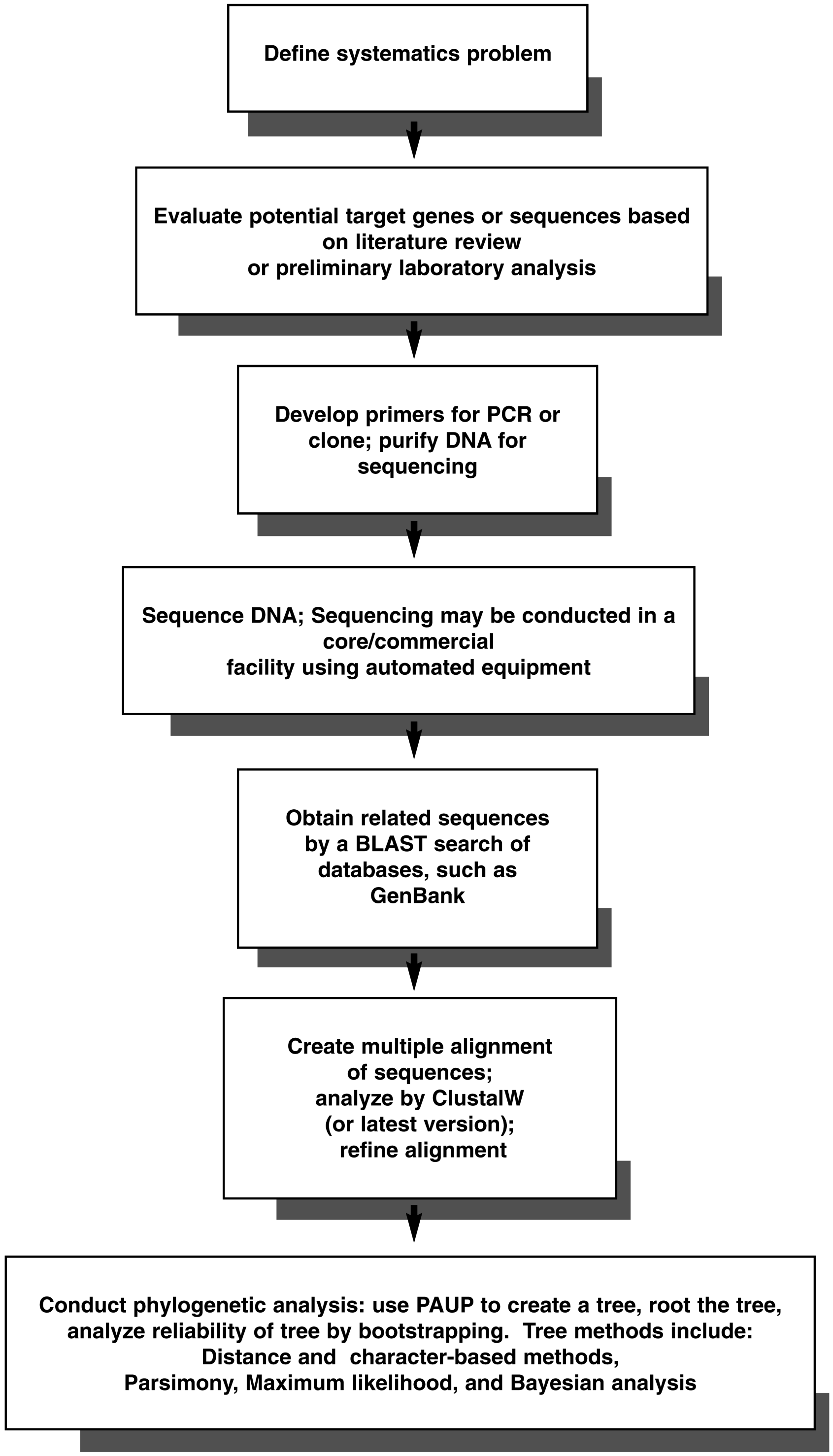


Figure 12.4. Steps in the phylogenetic analysis of DNA sequence data.

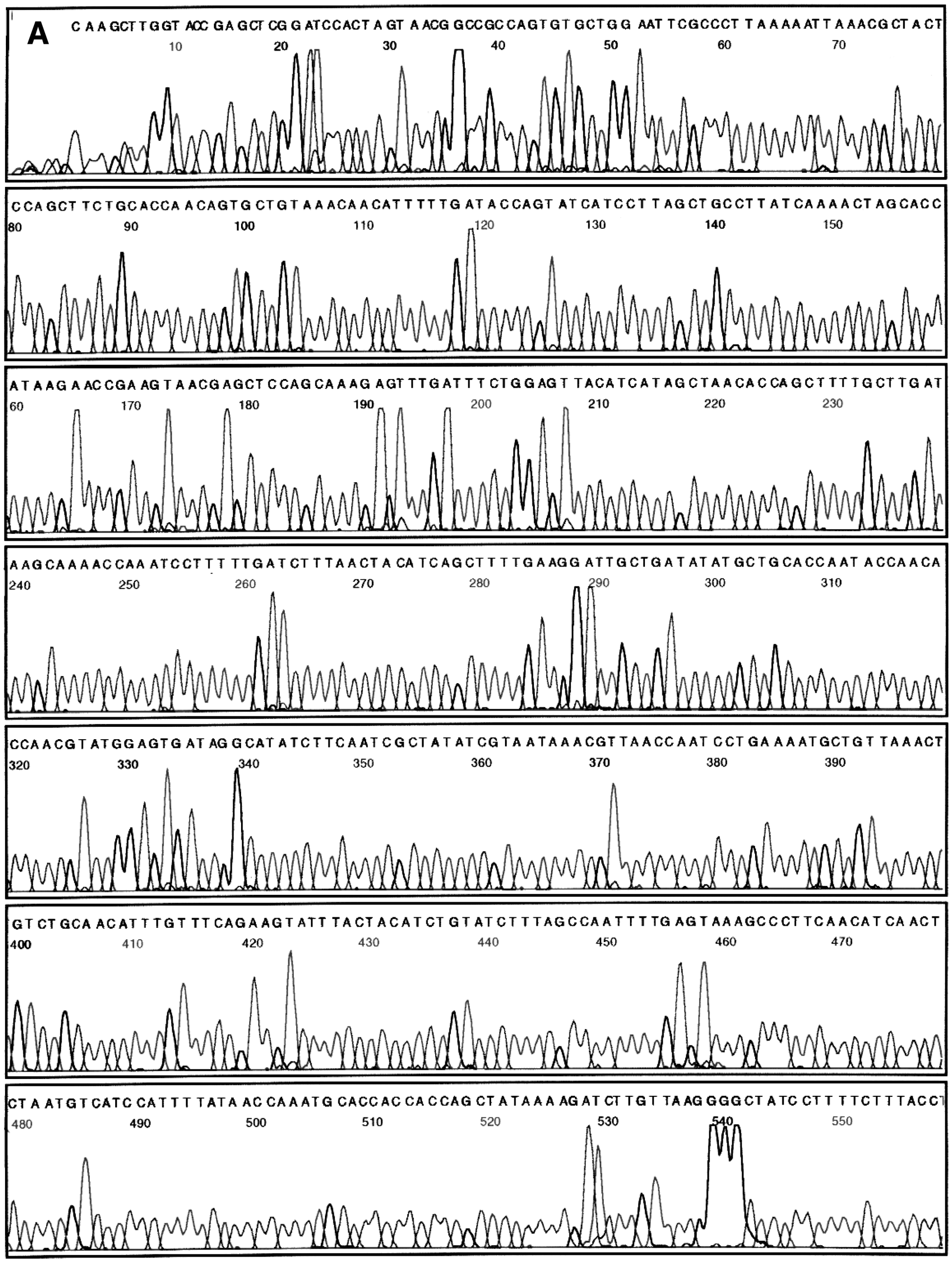


Figure 12.5. A) DNA sequence data are obtained from an automated DNA sequencing facility in this format. The different peaks representing A, T, C, and G are printed in different colors but are shown in black and white here. This sequence actually is the noncoding sequence and, thus, the sequence to be analyzed must be transformed into the complementary coding sequence. B) Once the coding sequence (● marked) is obtained, the sequence must be aligned by a computer program such as MacDNASIS and the coding strand analyzed. The open reading frame (ORF) is designated, and the beginning of the cloning vector sequence is shown at the end.

B **ORF ⟶ Coding Strand**

```
                     18                  27                  36                  45                  54
5' AGG TAA AGA AAA GGA TAG CCC CTT AAC AAG ATC TTT TAT AGC TGG TGG TGG TGC
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   Arg *** ... ... ... *** ... ... ... ... ... ... ... ... ... ... ... ...
 ●  Gly Lys Glu Lys Asp Ser Pro Leu Thr Arg Ser Phe Ile Ala Gly Gly Gly Ala
     Val Lys Lys Arg Ile Ala Pro *** ... ... ... ... *** ... ... ... ... ...

           63                  72                  81                  90                  99                 108
   ATT TGG TTA TAA AAT GGA TGA CAT TAG AGT TGA TGT TGA AGG GCT TTA CTC AAA
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... *** ... ... *** ... *** ... *** ... *** ... ... ... ... ...
 ●  Phe Gly Tyr Lys Met Asp Asp Ile Arg Val Asp Val Glu Gly Leu Tyr Ser Lys
     ... ... ... ... ... Met Thr Leu Glu Leu Met Leu Lys Gly Phe Thr Gln Asn

          117                 126                 135                 144                 153                 162
   ATT GGC TAA AGA TAC AGA TGT AGT AAA TAC TTC TGA AAC AAA TGT TGC AGA CAG
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... *** ... ... ... ... ... ... ... ... *** ... ... ... ... ... ...
 ●  Leu Ala Lys Asp Thr Asp Val Val Asn Thr Ser Glu Thr Asn Val Ala Asp Ser
     Trp Leu Lys Ile Gln Met *** *** ... ... ... ... ... Met Leu Gln Thr Val

          171                 180                 189                 198                 207                 216
   TTT AAC AGC ATT TTC AGG ATT GGT TAA CGT TTA TTA CGA TAT AGC GAT TGA AGA
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... ... ... ... ... ... *** ... ... ... ... ... ... ... *** ...
 ●  Leu Thr Ala Phe Ser Gly Leu Val Asn Val Tyr Tyr Asp Ile Ala Ile Glu Asp
     *** ... ... ... ... ... ... ... ... ... ... ... ... *** ... ... ... ...

          225                 234                 243                 252                 261                 270
   TAT GCC TAT CAC TCC ATA CGT TGG TGT TGG TAT TGG TGC AGC ATA TAT CAG CAA
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 ●  Met Pro Ile Thr Pro Tyr Val Gly Val Gly Ile Gly Ala Ala Tyr Ile Ser Asn
     ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...

          279                 288                 297                 306                 315                 324
   TCC TTC AAA AGC TGA TGT AGT TAA AGA TCA AAA AGG ATT TGG TTT TGC TTA TCA
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... ... *** ... ... *** ... ... ... ... ... ... ... ... ... ...
 ●  Pro Ser Lys Ala Asp Val Val Lys Asp Gln Lys Gly Phe Gly Phe Ala Tyr Gln
     ... ... ... ... Met *** ... ... ... ... ... ... ... ... ... ... ... ...

          333                 342                 351                 360                 369                 378
   AGC AAA AGC TGG TGT TAG CTA TGA TGT AAC TCC AGA AAT CAA ACT CTT TGC TGG
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... ... ... *** ... *** ... ... ... ... ... ... ... ... ... ...
 ●  Ala Lys Ala Gly Val Ser Tyr Asp Val Thr Pro Glu Ile Lys Leu Phe Ala Gly
     ... ... ... ... ... ... Met Met *** ... ... ... ... ... ... ... ... ...

          387                 396                 405                 414                 423                 432
   AGC TCG TTA CTT CGG TTC TTA TGG TGC TAG TTT TGA TAA GGC AGC TAA GGA TGA
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... ... ... ... ... ... ... *** ... *** *** ... ... *** ... ***
 ●  Ala Arg Tyr Phe Gly Ser Tyr Gly Ala Ser Phe Asp Lys Ala Ala Lys Asp Asp
     ... ... ... ... ... ... Met Val Leu Val Leu Ile Arg Gln Leu Arg Met Ile

          441                 450                 459                 468                 477                 486
   TAC TGG TAT CAA AAA TGT TGT TTA CAG CAC TGT TGG TGC AGA AGC TGG AGT AGC
   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 ●  Thr Gly Ile Lys Asn Val Val Tyr Ser Thr Val Gly Ala Glu Ala Gly Val Ala
     Leu Val Ser Lys Met Leu Phe Thr Ala Leu Leu Val Gln Lys Leu Glu *** ...

          495 | Vector Sequence ⟶ 513                 522                 531                 540
   GTT TAA TTT|TTA AGG GCG AAT TCC AGC ACA CTG GCG GCC GTT ACT AGT GGA TCC
   --- --- ---|--- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
   ... *** ...|... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
 ●  Phe Asn Phe *** ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     ... ... ... ... ... ... ... ... ... ... ... ... ... ... *** ... ... ...

          549                 558
   GAG CTC GGT ACC AAG CTT G 3'
   --- --- --- --- --- --- -
   ... ... ... ... ... ...
    ... ... ... ... ... ...
     ... ... ... ... ...
```

Figure 12.5. *continued*

12.6.5. Sequence Comparisons with BLAST

Once sequences have been obtained, the scientist usually wishes to compare them with other sequences in the databases (Figure 12.4). Genetic sequences are computerized and stored in three major databases: GenBank in the United States, the EMBL Data Library in Europe, and the DNA Data Bank of Japan. The amount of sequence information available for DNA has grown exponentially and continues to accumulate at an ever faster pace. The importance of submitting DNA sequence data to databases is recognized by many journals that require that sequences be submitted prior to, or simultaneous with, publication.

Sequences can be obtained from these databases using a computer program called **BLAST** (Figure 12.4). BLAST is one of the most widely used tools in phylogenetic analysis and is the Basic Local Alignment Search Tool. BLAST is used to search large databases of DNA (or amino acid) sequences, returning sequences that have regions of similarity to the sequence of interest provided by the user (query sequence) (Fortna and Gardiner 2001, Gibson and Muse 2002). In the BLAST program, the goal is to find regions in sequence pairs that have high levels of similarity. The results of a BLAST search orders the sequences and provides an e-value. The e-value is the number of hits with the same level of similarity that would be found by chance if there were no true matches in the database; thus, an e-value of 0.01 would occur once every 100 searches even when there is no true match in the database. BLAST searches can be run over the Web through the National Center for Biotechnology Information (NCBI), the European Biotechnology Institute (EBI), or the DNA database of Japan (DDBJ). Once the sequences have been obtained with which the data are to be compared, they need to be aligned.

12.6.6. Aligning Sequences

Sequences can be aligned either with other sequences obtained in the project or with sequences obtained from databases such as GenBank (Figure 12.4). Aligning the sequences usually involves computer analyses of the sequences using one of three major methods for comparing sequence similarity: matrix plots, global alignments, or local alignments (Hillis et al. 1990, 1996). Both alignment and phylogenetic inferences involve assumptions and subjective decisions (Hillis et al. 1990, 1996, Howe and Ward 1989, Gribskov and Devereaux 1991, Hall 2001). The alignments usually are made based on the assumption of parsimony. **Parsimony** dictates that an alignment of sequences is based on the minimal number of changes needed to transform one sequence into the other.

ClustalW is a commonly used computer program that aligns DNA (or amino acid) sequences in such a way as to maximize the number of residues that match by introducing gaps or spaces into one or the other sequence. These gaps are assumed to be due to insertions or deletions that occurred as the sequences diverged from a common ancestor over evolutionary time (Hall 2001).

12.6.7. Constructing Phylogenies

The primary methods of phylogeny construction are parsimony, distance, and likelihood, with many variants within each of these broad categories. The goal of all methods is to identify the relationships (topology of a tree) that are most congruent with the observed data. Many reviews of phylogenetic methods are available (Swofford and Olson 1990,

Swofford et al. 1996, Felsenstein 1988, 2002, Hillis et al. 1996, Pagel 1999, Shoemaker et al. 1999, Fox et al. 1999, Steel and Penny 2000, Huelsenbeck et al. 2001, Gibson and Muse 2002), and providing detailed procedures for constructing phylogenies is beyond the scope of this chapter. Only a brief outline of the different approaches is provided. Details of phylogenetic methods should be obtained from the reviews, books, and "how-to" manuals just cited. One should be aware that the methods for inferring phylogenetic relationships from molecular data continue to evolve. Phylogenetic analysis involves knowledge of statistics, computers, and mathematics, including calculus and matrix algebra; previous exposure to the theory of quantitative genetics is useful (Felsenstein 2002).

Inferring a phylogeny is an estimation procedure and is based on incomplete information. Any study of DNA sequences sampled from different species or different individuals in a population is likely to start with a phylogenetic analysis. Thus, phylogenetic analysis is becoming ever more common in biology. However, a novice will be frustrated by the fact that there are so many different approaches and different experts cannot agree.

The selection of one or more trees from among the set of possible phylogenies is based on one of two approaches: 1) defining a specific sequence of steps, an algorithm, for constructing the best tree, or 2) defining a criterion for comparing alternative phylogenies to one another and deciding whether they are equally good, or whether one is better. Some methods of phylogeny construction are based on different explicit evolutionary assumptions, while others are not.

Phylogenetic trees represent evolutionary pathways, and there is a difference between species trees and gene trees (Goldstein and Harvey 1999). Branches in a species tree join extant species to an ancestral species and represent the time since those species diverged. The data used to construct the tree often represent a single region of the genome of those species. A gene tree constructed from a short region of the genome may not be the same as the species tree. Two species may carry genes that diverged prior to the species split, or introgression or transposition may have resulted in genes having diverged after the species split.

Phylogenies are presented as rooted or unrooted trees. A **rooted tree** conveys the temporal ordering of the species or genes on a tree, but an **unrooted tree** reflects the distances between units with no notion of which was ancestral to which. Most of the analytical techniques result in an unrooted tree or unrooted phylogeny, one in which the earliest point in time is unidentified (Figure 12.3). In molecular phylogenies, **branch length** is the average number of nucleotide substitutions per site. If a branch length is 0.2, then on average the site has undergone 0.2 changes. Because a nucleotide changes or it doesn't, this average is based on 0 or 1 change.

Molecular data used to construct trees are either discrete characters or similarities (**distances**). Examples of discrete molecular characters include DNA sequences, allozyme frequencies, or restriction map data. Most methods assume independence and homology among discrete characters. Distance data specify a relationship between pairs of taxa or molecules. Sequence, restriction-map, and allozyme data must be transformed to produce distance data. Once data have been gathered and transformed into appropriate values, there are four broad categories of methods to estimate phylogeny. These include distance-matrix methods, maximum parsimony methods, and maximum likelihood methods, which are discussed in detail by Swofford and Olson (1990), Weir (1990), Hillis et al. (1996), Nei (1996), Huelsenbeck and Rannala (1997), Steel and Penny (2000), Hall (2001), and Whelan et al. (2001). A more recent addition to phylogenetic analysis involves Bayesian inference (Shoemaker et al. 1999, Huelsenbeck et al. 2001).

Distance matrix methods are based on the set of distances calculated between each pair of species and are the oldest family of phylogenetic reconstruction methods. The computations are relatively simple, and the quality of the resulting tree depends on the quality of the distance measure. Using distances to group the taxonomic units into a phenetic grouping usually employs clustering.

Several methods of clustering can be used, but the most widely used is called **UPGMA** (Unweighted Pair-Group Method using an Arithmetic average). It defines the intercluster distance as the average of all the pairwise distances for members of two clusters. The results of the clustering can be presented in a dendrogram, in which the branch points are placed midway between two sequences or clusters. The distance between a pair of sequences is the sum of the branch lengths. The UPGMA often is used for distance matrices, and it generally performs well when the mutation rates are the same along all branches of the tree. However, the assumption of nearly equal mutation rates (or that a molecular clock is operating) is crucial.

For situations in which the assumptions of the molecular clock are inappropriate, the Fitch–Margoliash algorithm can be used (Weir 1990). If information for an outgroup is available, the resultant tree can be rooted. The Fitch–Margoliash method allows for the possibility that the tree found is incorrect and recommends that other trees be compared based on a measure of goodness of fit. The best tree will have the smallest percentage standard deviation. The Fitch–Margoliash and UPGMA methods should result in very similar trees if a molecular clock is operating.

Maximum parsimony methods focus on the character values observed for each species, rather than working with the distances between sequences that summarize differences between character values. These methods *minimize* the numbers of changes in sequences between species over the tree, usually making the assumption that there have been approximately constant rates of change. Branch lengths usually are not obtained. Maximum parsimony is widely used and works well when change is rare or branches are short (Pagel 1999). Parsimony methods can work poorly when rates of character evolution are high, and the phylogeny includes some long branches because it tends to underestimate the amount of change in long branches. In some circumstances, maximum-likelihood and parsimony methods can provide equivalent results.

For each possible tree, the sequences at each node are inferred to be those that require the least number of changes to give each of the two sequences of the immediate descendants. The total number of changes required over the whole tree is found, and the tree with the minimum number of changes is the most parsimonious. Parsimony methods assume that genetic changes are improbable. However, if there are large amounts of change, parsimony methods can yield estimated trees that are inaccurate (Swofford and Olson 1990). Stewart (1993) pointed out that parsimony analysis can be problematic for two general reasons: 1) the shortest tree is not found and 2) the shortest tree is not the correct phylogeny. Failure to find the shortest tree can occur if too many taxa or too few informative data are used.

Likelihood methods of analyzing DNA sequence data rely on genetic models and can provide a basis for statistical inference. Likelihood is an amount proportional to the probability of observing the data, given a model. Likelihood methods are more difficult to compute (Weir 1990). Maximum likelihood methods of tree construction assume the form of the tree and then choose the branch length to maximize the likelihood of the data given that tree. These likelihoods then are compared over different possible trees, and the tree with the greatest likelihood is considered to be the best estimate.

Unfortunately, the number of possible trees increases very rapidly as the number of taxa under consideration increases. Thus, if three species are being compared, the number of possible unrooted trees is one; with four species it is three trees, with six species it is 105 trees, and with eight species, it is 10,395 trees. Maximum likelihood methods provide consistent estimates of branch lengths, indicating that the estimates approach the true values as the amount of data increases. To estimate the likelihood that a particular tree estimate is the true tree, **bootstrapping** techniques can be employed. Bootstrapping involves repeated sampling, with replacement, of artificial data sets to produce an estimate of the variance. The name of this statistical method was derived from the term "to pull yourself up by your bootstraps," and the method allows statistical distributions to be generated from very few data.

Methods for analyzing molecular data are still undergoing development, because none of the techniques currently available is fully satisfactory. The immense amount of DNA sequence data that is becoming available makes it difficult to use maximum likelihood methods unless very powerful computers are used. Maximum likelihood algorithms have been developed to build trees from pairwise distances, but they employ only a summary of the data and information is thus lost. Parsimony methods are fast, but may be appropriate only for very slow rates of evolutionary change.

12.6.7.1. Bayesian Methods

Another approach to analyzing evolutionary processes and phylogeny is Bayesian inference (Shoemaker et al. 1999, Huelsenbeck et al. 2001). Bayesian inference uses the same models of evolution as many other methods and can be used to infer phylogeny, evaluate uncertainty in phylogenies, detect selection, compare trees, evaluate divergence times, and test the molecular clock (Huelsenbeck et al. 2001).

Bayesian inference of phylogeny is based on a quantity called the "posterior probability of a tree" and uses Bayes's theorem:

$$\mathrm{Pr[Tree\ I\ Data]} = \frac{\mathrm{Pr[Data\ I\ Tree]} \times \mathrm{Pr[Tree]}}{\mathrm{Pr[Data]}}$$

In this theorem the vertical bar should be read as "given" and is used to

> combine the prior probability of a phylogeny (Pr[Tree]) with the likelihood (Pr[Data I Tree]) to produce a posterior probability distribution on trees (Pr[Tree I Data]). The posterior probability of a tree is the probability that the tree is correct. Inferences about the history of the group are then based on the posterior probability of trees and the tree with the highest posterior probability might be chosen as the best estimate of phylogeny (Huelsenbeck et al. 2001).

The likelihood is calculated under one of a number of standard Markov models of character evolution. A Markov process is a mathematical model of infrequent changes of discrete states (nucleotides or amino acids) over time, in which future events occur by chance.

Advocates of the Bayesian approach note that phylogenetic analysis can be difficult because a large number of trees potentially could describe the relationships of a group of species. Evaluating which of these trees are the best approximation of the "true" tree can be difficult when rates of DNA substitution are high; multiple substitutions at a site can make it difficult to resolve true relationships, producing the "wrong tree." Methods that explicitly deal with multiple substitutions can overcome the statistical problems, but the

most powerful methods (maximum likelihood) can be used only on relatively small data sets and many of the faster methods do not take advantage of the information of all the data contained in the DNA sequences.

Bayesian inference makes it possible to analyze large data sets more easily. Instead of searching for the optimal tree, trees are sampled according to their posterior probabilities. Once such a sample is available, features that are common among these trees can be discerned and a consensus tree can be constructed. "This is roughly equivalent to performing a maximum likelihood analysis with bootstrap resampling, but much faster" (Huelsenbeck et al. 2001).

Shoemaker et al. (1999) noted that a "common criticism of the Bayesian approach is that the choice of the prior distribution is too subjective." Thus, researchers using the same data could reach different conclusions if they used different prior distributions. Furthermore, implementation of Bayesian methods can be "very complex." Bayesian methods may be especially useful for analyzing complex evolutionary models (including horizontal gene transfer) and accommodating phylogenetic uncertainty.

12.6.8. Artifacts

Phylogenies of animal phyla constructed using 18S rRNA sequences may not be as accurate as originally thought (Maley and Marshall 1998, Lee 1999, Philippe and Forterre 1999, Philippe et al. 2000). For example, a phylogenetic analysis of the entire protein-coding mitochondrial genome of chordates and selected invertebrates yielded strong support for an incorrect tree using parsimony, distance, and likelihood methods. The basis of the incorrect trees was found to be due to three hydrophobic amino acids; when these sites were ignored, the tree became compatible with known relationships. Apparently these amino acids have undergone concerted evolution which has obscured the underlying historical phylogenetic signal. Lee (1999) concluded that both "morphological and molecular systematics might have more in common than previously assumed." By recognizing these possible difficulties, it is possible to address the problem.

Inaccuracies may occur for a variety of reasons (Adoutte et al. 2000). Alignments of corresponding sequences must be carried out carefully. If unambiguous alignments of sequences cannot be obtained, different relationships may be obtained. Poor alignments may result in a lack of strong statistical support for a particular tree. Another factor that affects DNA phylogenies is the species chosen to represent each group. Use of different species can result in different trees. Increasing the number of species analyzed will help resolve this problem, but the increased number of species increases the computational time required to find the best tree to represent the relationships. For example, if five species are studied there are just 15 possible unrooted trees, but with 50 species there are 3×10^{74} potential trees to analyze.

Inaccurate trees may occur because of a phenomenon called **long branch attraction**. When long branches on evolutionary trees are in close proximity to short branches, maximum parsimony will recover the wrong tree because the long branches tend to group together or "attract each other."

12.6.9. Software Packages

A wide array of computer software packages for phylogenetic analyses are available and supported (Swofford and Olsen 1990, Eernisse 1998). Software evolves rapidly, but

information is provided at the end of the chapter so that sources can be contacted for recent information and updated versions. The computer packages make it easy to conduct analyses using multiple methods. If the results are compared and there is concordance among the analyses, a particular tree is more likely to be correct (Caterino et al. 2000).

MrBayes is available from http://brahms.biology.rchester.edu/software.html if you wish to carry out a Bayesian analysis of your data. **PHYLIP** is the Phylogeny Inference Package, available from Joseph Felsenstein, Department of Genetics, University of Washington, Seattle. It is a collection of about 30 independent programs implementing maximum likelihood, parsimony, compatibility, distance, and invariant methods. Some of the programs provide bootstrap methods for estimating confidence limits. **PAUP** is Phylogenetic Analysis Using Parsimony, available commercially (http://paup.csit.fsu.edu). PAUP performs parsimony analysis under a variety of models, and bootstrapping routines are available. Also available is BIOSYS-2, which includes cluster analysis and distance Wagner routines for gene frequency data. **Hennig86** is a small, fast, and effective program for parsimony analysis under the Wagner and Fitch models. **MacClade**, written by W. P. Maddison and D. R. Maddison and distributed by Sinauer Associates (Sunderland, MA 01375), is useful in the analysis of character evolution and the testing of phylogenetic hypotheses under the same parsimony models described for PAUP plus additional ones.

12.7. The Universal Tree of Life

12.7.1. Two Domains

Until rather recently, the traditional view was that life is divided into "animals" and "plants"; the study of bacteria and fungi often took place in departments of botany. Later, it was realized that organisms can be divided into prokaryotes or eukaryotes. Under this grouping, it became clear that fungi are not plants (molecular data indicate they are actually more closely related to animals). All single-celled eukaryotes initially were placed into the phylum Protista, but this was found to be a heterogeneous group consisting of algae (formerly plants), protozoa (formerly animals), water molds (formerly fungi), and others (Mayr 1998).

12.7.2. Three Domains

A more recent view is that there are three primary "domains" of life (Archaeabacteria, Eubacteria, and Eukaryota) (Woese et al. 1990, Doolittle 1999, Woese 2000).

"Bacteria" were separated into two groups on the basis of variation in the small-subunit rRNA: the traditional bacteria (called Eubacteria) and a previously unrecognized group (Archaeabacteria), which contain members that inhabit extreme environments such as hot springs, sulfur springs, and deep vents (Woese 1987). Woese (1987) considered the Archaeabacteria might have been the first organisms on earth because they inhabit such extreme environments, although now we know Archaeabacteria are more widespread than this. However, separation of the Eu- and Archaeabacteria into two domains of a rank equal to that of the eukaryotes was justified because they were assumed to have evolved independently from a precursor group (Figure 12.6). The Archaeabacteria were considered to be as different from the Eubacteria as the Eubacteria were from the eukaryotes on a molecular basis.

Separation of life into three domains remains controversial (Cavalier-Smith 1998, Mayr 1998). Mayr (1998) argues that the three-domain arrangement is unjustified because the

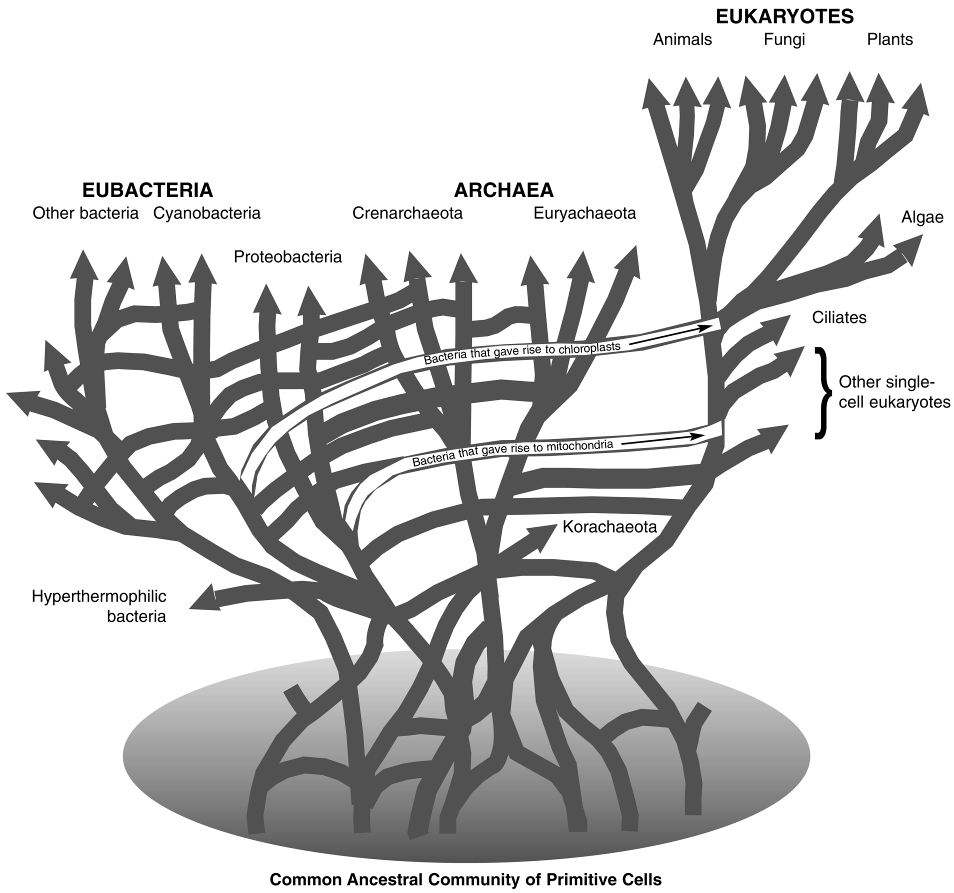


Figure 12.6. Molecular phylogenies support the concept of a "web" of life, in which gene exchange and horizontal gene transfer have had significant effects on the evolution of the Archaea, Eubacteria, and Eukaryota. Evolution of life has involved multiple events of horizontal gene transfer between the domains, including incorporation of Eubacteria that gave rise to chloroplasts and mitochondria, as well as evolution of vertically transmitted archaeal genes. A "linear" view of the evolution of life had to be modified once genome analysis indicated the relationships between the domains were more complex. (Modified from Doolittle 2000.)

number of groups known for Archaeabacteria (175) and Eubacteria (10,000) is far exceeded by the number of eukaryotic species (~30 million), and the *phenotypic* diversity is orders of magnitude less for the Archaeabacteria and Eubacteria than for the eukaryotes.

12.7.3. Origin of Eukaryotes

Molecular, geological, and paleontological evidence suggests that eukaryotes originated around 2 billion years ago (Katz 1998, 1999, Figure 12.6). Under one model, eukaryotic cells acquired mitochondria only after the divergence of several extant eukaryotic lineages. The rRNA genealogy and analyses of ancient gene duplications of protein-coding genes support the sister status of Archaea and Eukaryotes (Katz 1998). However, this view had to be modified after complete genome sequences were obtained from a variety

of organisms. There is a great likelihood that eukaryotic genomes are **chimeric**, derived from both archaeabacterial and eubacterial lineages (Golding and Gupta 1995, Rivera et al. 1998, Lang et al. 1999, Roger 1999).

Chimeric genomes could have developed only if lateral transmission of genes (or genomes) took place across species boundaries (Katz 1999). The genomic evidence suggests that informational genes (genes involved in transcription, translation, and related processes) are typically members of large complex systems and are less likely to be transferred than operational genes (genes involved in "housekeeping") (Jain et al. 1999).

Doolittle (1998) proposes that gene swapping (leading to chimerism) among early organisms may have occurred when primitive eukaryotes picked up genes from their food, and he suggested that "You are what you eat." According to this scenario, the ancestor of eukaryotes was archaeal, but many eukaryotic nuclear genes today are of eubacterial origin because horizontal transfer occurred. Some of the eubacterial genes could have moved horizontally from mitochondria (probably originally a proteobacterium), but these genes of mitochondrial origin were relatively few and limited to the proteins that are reimported into the mitochondria. Doolittle (2000) speculates that the eubacterial genes with other functions could have moved into the eukaryotic genome when phagocytic unicellular eukaryotes fed on α-proteobacterium. DNA from these food bacteria would have moved repeatedly into the nuclear genome. Doolittle (1998) argues that

> all genes that can be replaced by food-derived [eubacterial] genes will be, in the fullness of time. We should not think of such gene replacement as idiosyncratic or exceptional, but as the normal course. It is, instead, the persistence of some genes of archaeal ancestry that requires special explanation.

A consequence of lateral (horizontal) gene transfer is that phylogenetic analyses of different genes can result in conflicting phylogenies, causing confusion (Katz 1998, Bushman 2002). Evidence from sequences of 66 protein-coding genes from members of all three domains suggests that some eukaryotic genes are more similar to archaeal genes, while others appear to share ancestry with eubacterial genes (Brown and Doolittle 1997, Katz 1998). Genes involved in the genetic machinery of the cell are shared by the Archaea and eukaryotes, while genes that regulate metabolic processes are shared by Eubacteria and eukaryotes. Thus, these analyses "challenge the traditional view that vertical transmission of genetic material from one generation to the next is the predominant force in evolution" (Katz 1998).

The origin of eukaryotes continues to be studied and debated. Key characters involved in the emergence of eukaryotes include the presence of a nucleus, microtubules, mitochondria, and a chimeric genome (Katz 1998). How all these parts were assembled remains controversial; some hypothesize there was a single endosymbiosis event, and others believe that there were two or more endosymbiosis events. It is possible that the original event that gave rise to mitochondria occurred in the ancestor of all extant eukaryotes, which could explain both the chimeric nuclear genome and the origin of mitochondria.

The origins of the nucleus and microtubules are less well understood; hypotheses on the origin of the nucleus include: 1) the nucleus is derived from a nucleoid structure found in Archaea; 2) the nucleus evolved through invagination of membranes within the lineage that gave rise to eukaryotes; 3) the nucleus resulted from the engulfment of one organism by another. Katz (1998) argues that distinguishing among these hypotheses is likely to be difficult and concludes by noting that

> individual gene genealogies alone cannot provide a "tree of life." Instead, these genealogies spin a tangled web of the history of genes within organisms. The next steps are to develop more

sophisticated methods to interpret conflict among multiple gene genealogies, and to augment molecular data with studies of the cell biology of early diverging eukaryotes.

Dacks and Doolittle (2001) suggest that more data are needed to understand the origin of eukaryotes: 1) identify the surviving direct descendants of primitively simple eukaryotes that must have lived in ancient anaerobic habitats and study their genomic, physiological, and structural diversity; and 2) construct phylogenies using sequences from more different genes to resolve questions (for example, an insertion in the EF1 gene was used as evidence to conclude animals and fungi are more closely related than are fungi and plants).

Genome duplication is thought to be an important component of the evolution of eukaryote genomes (Ohno 1970, Wagner 1998, Sankoff 2001). Genome or gene duplication is thought to be the predominant method by which new gene functions can evolve, despite the fact that the vast majority of duplicate genes are expected to become pseudogenes through mutations. The loss of a duplicate gene is expected, because as long as one gene functions normally the other can accumulate deleterious mutations.

Many protein-coding genes belong to multigene families, which could have evolved by gene duplication (Friedman and Hughes 2001). Analysis of the complete genomes of *D. melanogaster*, the nematode *Caenorhabditis elegans*, and the yeast *Saccharomyces cerevisiae* showed that duplication of genomic blocks has occurred, although the duplications did not all occur at the same time (Friedman and Hughes 2001). Some blocks could have been due to an ancient polyploidization event; others are more recent and could have involved duplications of chromosome segments. Some duplicated blocks in the yeast genome are associated with transposable elements that could have been the cause of the duplication events.

12.8. The Fossil Record of Arthropods

Insects have a relatively extensive fossil record, with 1263 families of fossil insects known (Labandeira and Sepkoski 1993). Labandeira and Sepkoski (1993) found 472 references on fossils covering 1263 insect families with all of the 30 commonly recognized extant orders of insects represented as fossils. Although only a few fossil insects (such as Collembola) are known from the lower Devonian, a massive radiation began sometime during the early Carboniferous, more than 325 million years ago, and the pterygotes radiated into stem groups of all major lineages, including ephemeroids, odonatoids, plecopteroids, orthopteroids, blattoids, hemipteroids, and endopterygotes. Insects continued to increase in diversity during the late Carboniferous and middle Permian (Table 12.4).

Insects are highly diverse and ancient arthropods. The Crustacea are considered the sister group of the Tracheata (=Myriapoda + Hexapoda or Insecta). Relatively advanced Crustacea are found in the Cambrian (600 mya), so it is assumed that tracheates were present by this time as well (Kukalova-Peck 1991). Labandeira et al. (1988) showed that a bristletail (Archaeognatha) from the Early Devonian resembles modern archaeognathans. Arthropods have apparently been found on land since Devonian times (Table 12.4). Two Collembola species were found in the lower Devonian (400 mya) that resemble recent extant Isotomidae and Neanuridae, suggesting that terrestrial arthropods already had radiated in the Ordovician (ca. 500 mya).

A number of extinct and extant orders of primitive insects have been found in a diverse late Paleozoic fauna (Table 12.4). During the Carboniferous (which began 360 mya)

Table 12.4. Geological Time Scale in Millions of Years and Types of Fossil Insects Found

Era	Period	Epoch	Began (mya)	*Extinct* and extant fossil insect orders first found
Cenozoic	Quaternary	Recent		Protura, Zoraptera, and Phthiraptera first appeared in fossil record.
		Pleistocene	1.6	
	Tertiary	Pliocene	5	
		Miocene	25	
		Oligocene	35	
		Eocene	60	Mantodea first appeared in fossil record.
		Paleocene	65	
Mesozoic	Cretaceous		145	Isoptera first appeared in fossil record.
	Jurassic		210	Dermaptera first appeared in fossil record.
	Triassic		245	Odonata, *Titanoptera*, Grylloblattodea, Trichoptera, Lepidoptera, and Hymenoptera first appeared in the fossil record.
Paleozoic	Permian		285	*Permothemistida*, Plecoptera, Embioptera, *Protelytroptera*, *Glosselytrodea*, Psocoptera, Thysanoptera, Hemiptera, *Antliophora*, Mecoptera, Diptera, *Amphiesmenoptera*, Neuroptera, Megaloptera, and Coleoptera first appeared in the fossil record.
	Carboniferous		360	Pterygotes radiated into stem groups of all major lineages, with seven surviving to modern times (ephemeroids, odonatoids, plecopteroids, orthopteroids, blattoids, hemipteroids, and endopterygotes). Diplura, *Monura*, Thysanura, *Diaphanopterodea*, *Megasecoptera*, *Permothemistida*, *Protodonata*, *Paraplecoptera*, Orthoptera, Blattodea, *Caloneurodea*, *Blattinopsodea*, and *Miomoptera* were present.
	Devonian		400	Collembola *(Rhyniella praecursor)* and *Archaeognatha*
	Silurian		440	Mites, opilionids, scorpions, pseudoscorpions, centipedes, spiders found in pre-Devonian strata
	Ordovician		500	
	Cambrian		600	

Modified from Kukalova-Peck (1991).

a diverse array of extinct and extant (underlined) insects were present, including the: Diplura, Monura, Diaphanopterodea, Palaeodictyoptera, Megasecoptera, Permothemistida, Ephemoptera (mayflies), Protodonata, Paraplecoptera, Plecoptera (stoneflies), Orthoptera (grasshoppers and crickets), Blattodea (cockroaches), Caloneurodea, Blattinopsodea, and Miomoptera.

During the Permian (which began 285 mya), additional extinct and extant insect groups are found in the fossil record, including Plecoptera (stoneflies), Embioptera (web spinners), Protelytroptera, Glosselytrodea, Thysanoptera (thrips), Hemiptera (bugs and leafhoppers), Antliophora, Mecoptera (scorpion flies), Diptera (true flies), Amphiesmenoptera, Neuroptera (lacewings, antlions), Megaloptera (dobsonflies), and Coleoptera (beetles).

By the Triassic (245 mya), nearly all modern orders of insects are found in the fossil record, including Lepidoptera (butterflies and moths), Trichoptera (caddisflies), and Hymenoptera (bees and wasps). By the Jurassic (210 mya), many recent families are present.

Tertiary insects (65 mya) are essentially modern and include genera nearly indistinguishable from living fauna.

Labandeira and Sepkoski (1993) suggest that the great diversity of insects was achieved by low extinction rates rather than by high origination rates. The great radiation of modern insects began 245 million years ago and was not accelerated by the expansion of the angiosperm plants during the Cretaceous period.

There are more than 700,000 living species of living insects classified in at least 29 orders and more than 750 families, with the orders Coleoptera (>300,000 named species), Lepidoptera (>120,000 species), Hymenoptera (>120,000 species), and Diptera (>150,000 species) containing the most species. Insects are diverse, numerous, and ancient (Daly et al. 1998). An understanding of their systematics and phylogeny requires the combined use of the fossil record, traditional morphological data, and molecular methods.

12.9. Molecular Analyses of Arthropod Phylogeny

12.9.1. Ribosomal RNA

The origin and phylogeny of the Arthropoda have been analyzed using ribosomal sequences (Field et al. 1988, Turbeville et al. 1991). The slowly evolving core segments of rRNA allow the reconstruction of phylogenies of phyla and kingdoms. To study the phylogenetic relationship at the family level, domains that change more rapidly must be analyzed, such as a 324-bp sequence from the second expansion segment of the 28S gene.

The molecular phylogeny of the Metazoa was investigated by sequencing the 18S rRNA gene (Field et al. 1988, Adoutte et al. 2000). The branching patterns obtained by traditional phylogenies based on morphology and embryology can be compared with branching patterns derived from rRNA data (Figure 12.7). A traditional phylogeny indicates that arthropods are not close relatives of the annelids, suggesting an early divergence of arthropods from other metameric lineages.

In the tree based on the 18S rRNA sequences, chelicerates are represented by the horseshoe crab *Limulus*, crustaceans by the brine shrimp *Artemia*, Uniramia by *Drosophila*, and the millipede by *Spirobolus*. Field et al. (1988) note that, although the number of arthropod species they sampled is limited, the results support the hypothesis that arthropods represent a coherent phylum of single origin, rather than a polyphyletic group with several distinct annelid ancestors. However, Field et al. (1988) sampled relatively few arthropods and were unable to resolve relationships within the Arthropoda.

A study to resolve relationships within the Arthropoda compared partial 18S rRNA sequences of five chelicerate arthropods, plus a crustacean, myriapod, insect, chordate, echinoderm, annelid, and platyhelminth (Turbeville et al. 1991). The sequence data were analyzed using a maximum-parsimony method, an evolutionary-distance method, and the evolutionary-parsimony method. The results generated by maximum-parsimony and distance methods support monophyly of the Arthropoda and monophyly of the Chelicerata within the Arthropoda, which are congruent with phylogenies based on cladistic analyses of morphological characters.

As shown in Figure 12.7, the rRNA-based molecular phylogenies show the Bilaterians as a monophyletic group clearly separated from sponges, cnidarians, and ctenophores (Adoutte et al. 2000). The clade uniting annelids and arthropods on the basis of segmentation of the body trunk is traditional, but the rRNA tree places annelids and mollusks

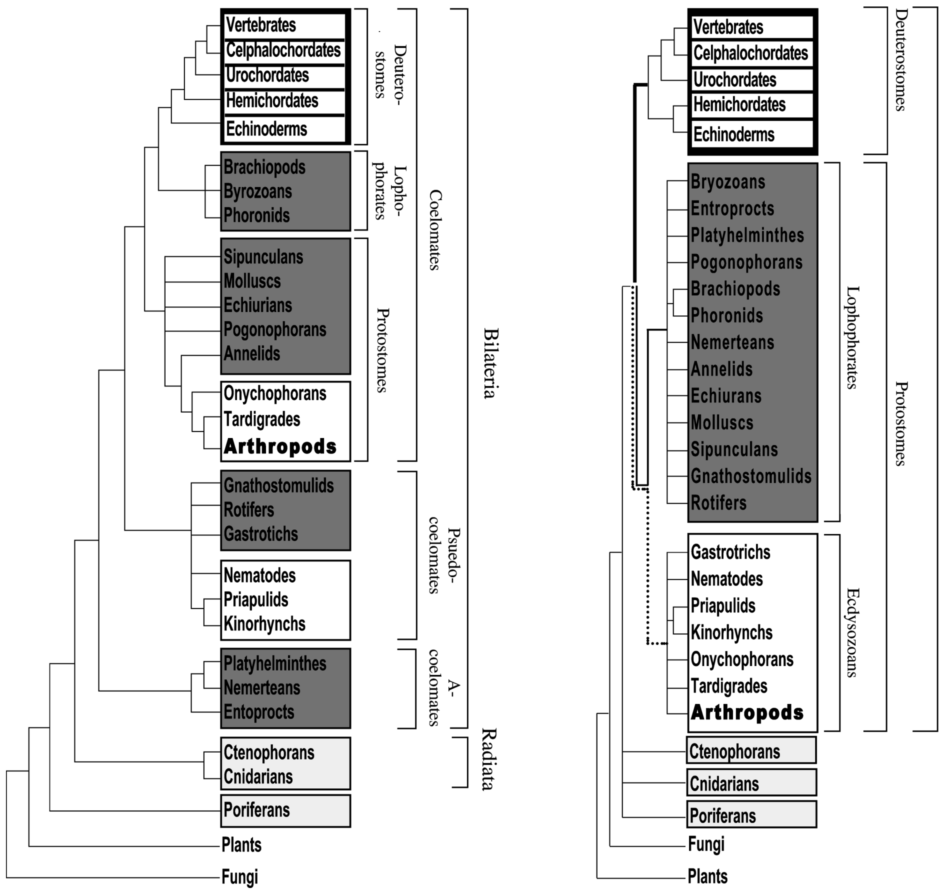


Figure 12.7. Phylogenies of the Metazoa compared: A) A traditional phylogeny is based on morphology and embryology, with organization of the Bilateria into Coelomates, Pseudocoelomates, and Acoelomates. Using this scheme, Arthropoda are associated with Onychophora and Tardigrada and the Nematoda are Acoelomates. B) A molecular-based phylogeny places the Arthropoda within the Ecdysozoans among the Bilateria, which includes Tardigrada, Onychophora, and Nematoda. (Redrawn from Adoutte et al. 2000.)

and other unsegmented phyla in a Lophotrochozoan clade and places the Arthropoda in an Ecdysozoan clade with Nematoda, Onychophora, and Tardigrada. This is consistent with the fact that annelids share with mollusks a typical mode of spiral egg cleavage, followed by the formation of a trochophore larva. One of the most interesting changes is that nematodes are now considered a sister group of arthropods but had been placed outside the Bilaterians because of a long-branch attraction artifact. The nematodes, arthropods, and related phyla all have a molting cuticle; hence the name Ecdysozoa (Figure 12.7). Analysis of β-thymosin sequences also support the arthropod-nematode clade (Manuel et al. 2000).

12.9.2. *Hox* Genes and Arthropod Phylogeny

Traditional taxonomy and molecular-based phylogenies of the Arthropoda may appear to be incongruent because the morphological characters used to build them have undergone

convergence, particularly when insects and myriapods are considered (Cook et al. 2001). Molecular trees have been difficult to build because arthropod diversification probably was rapid and ancient, with major arthropod groups already present in the early Cambrian approximately 520 mya. Thus, the few traits that provide relevant phylogenetic information may be misleading because there have been more than 500 million years of evolutionary change. Analysis of *Hox* genes provides an interesting window on the evolutionary history of arthropods.

Hox genes are an ancient family that regulate differentiation of anterior/posterior axes of the arthropod body. *Hox* genes contain a homeodomain region and are important in organizing specialized sections such as head, thorax, and abdomen. *Hox* genes have been present since the early divergence of Bilaterians. Cook et al. (2001) sequenced *Hox* genes from an oribatid mite (Chelicerata), a symphylan (Myriapoda), pauropods (Myriapoda), a branchiopod (Crustacea), and a grasshopper (Insecta) and compared them to other arthropod *Hox* sequences. The resulting phylogenetic analysis supported a "hexapod/crustacean clade . . . to the exclusion of myriapods" and the data suggest that Myriapoda are more closely allied to the Chelicerata than to the Insecta /Crustacea clade (Cook et al. 2001).

The Insecta and Myriapoda have long been considered to be most closely related among the Arthropoda, primarily because they share trachea (Tracheata) and Malpighian tubules and lack second antennae. The *Hox* data suggest that Malpighian tubules and tracheae of Insecta and Myriapoda may have evolved convergently, while their secondary antennae were lost convergently. Cook et al. (2001) also predicted that "insects must derive not from some homonomous myriapod-like body but rather from an already tagmatized crustacean, with very different implications for the evolution of segmentation." The molecular data suggest that mandibles might have been present in the common arthropod ancestor and were lost in the chelicerates. Alternatively, Cook et al. (2001) conclude that evolution of mandibles in myriapods and in the crustacean/insect clade was convergent and the results reinforce the "conclusion that the morphological features traditionally used to infer relationships among the arthropod subphyla make a poor phylogenetic data set. At this depth in the tree, convergence and stochastic change overwhelm whatever phylogenetic signal they contain."

12.9.3. *Hox* Genes and Evolution of Arthropod Appendages

The duplication and diversity of *Hox* genes, changes in their regulation, and changes in the regulation of genes targeted by *Hox* genes are important in understanding the evolution of arthropods (Doolittle 1999, Levine 2002, Ronshaugen et al. 2002). The success of arthropods is correlated with their modular body plan—a series of repeating segments that can be modified to contain legs, wings, antennae, mouthparts, or genital structures. Insects have six legs, two on each of the three thoracic segments, whereas crustaceans have a variable number of swimming appendages; some crustaceans have limbs on every segment in both the thorax and abdomen.

Galant and Carroll (2002) identified a transcriptional repression domain in the carboxy-terminal region of the *Drosophila* Ultrabithorax protein that is highly conserved among insects, but is absent from this gene in other arthropods and onychophorans. Galant and Carroll (2002) speculate that this domain could have facilitated the diversification of posterior thoracic and anterior abdominal segments that is characteristic of modern insects. Ronshaugen et al. (2002) provide evidence that suppression of abdominal limbs in insects depends on functional changes in *Ultrabithorax*. Thus, mutations in a homeotic gene apparently resulted in significant morphological changes approximately 400 million years ago

when six-legged insects limited the production of limbs to the thorax and diverged from crustacean-like arthropod ancestors that had limbs on the abdomen as well (Ronshaugen et al. 2002).

12.9.4. Other Molecular Phylogenies of Insects

Other phylogenies of arthropods and insects have been conducted using several highly conserved genes. For example, Regier and Shultz (1997) examined the evolution of the arthropods using sequences from *elongation factor-1* and the largest subunit of RNA polymerase II, and Burmester et al. (1998) used *hexamerin* sequences. Wheeler et al. (2001) combined morphological characters and 18S and 28S rRNA sequences and compared the results of trees of insect orders produced by the morphological data with trees produced by molecular data.

12.9.5. Congruence between Morphology- and Molecular-Based Trees

Unfortunately, the fossil record is often inadequate, so inferences made about lineages of organisms are based on what the scientist can observe and measure. Many scientists are concerned about using a single method, such as sequence data or morphological traits, to deduce evolutionary patterns. The possibility exists that inferences concerning phylogenetic relationships based on molecular data may not reflect accurately the historical relationship of the taxa from which the data were obtained, producing a "gene tree/species tree" problem. For example, Powell (1991) pointed out that molecular studies on the *Drosophila pseudoobscura* group can lead to conclusions of monophyly, paraphyly, and polyphyly, depending upon which data are used to construct the trees. The different sets of data used are presumed to be accurate and thus neither tree is "wrong," but reflects different aspects of the history of the same taxa when different data are considered. Morphological and molecular data can lead to different conclusions in some cases, but produce congruent results in others.

12.10. Molecular Evolution and Speciation

12.10.1. Species Concepts

One of the central questions of biology is how a continuous process of evolution can produce "species" (Coyne 1992, Rice and Hostert 1993, Hollocher 1998). At least three different views of "species" are used (O'Hara 1994). The **biological species concept** indicates, "Species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups" (Mayr 1970). Reproductive isolation is achieved by **prezygotic isolating factors** (mating discrimination, different habitat preferences) and **postzygotic isolating factors** (hybrid inviability, sterility). Reproductive isolation, in concert with selection and genetic drift, creates and expands the morphological differences between species living in the same area.

Physical isolation (**allopatry**) leads inevitably to evolutionary change through natural selection or drift, and pre- or postmating reproductive isolation mechanisms evolve as a by-product of the genetic changes. Any resultant hybrid inviability could be due to the development of divergent developmental systems. Reproductive isolation may be increased if incompletely isolated populations become **sympatric** (live together in the same area) so that selection would fix the alleles that reduce interspecific mating. The process

of increasing isolation is called "reinforcement," but how often this process occurs is debated. Likewise, the extent of sympatric speciation, in which reproductive isolation occurs without geographic isolation, remains controversial, although the sympatric host races of the tephritid *Rhagoletis pomonella* represent one well-documented example (Feder et al. 1988, 1997). The biological species concept, in the view of some, overemphasizes reproductive isolation between populations.

The **evolutionary species concept** emphasizes the continuity of populations through time. Thus an "evolutionary species is a lineage (ancestral-descendant sequence of populations) evolving separately from others with its own unitary evolutionary role and tendencies" (Simpson 1961). This definition focuses more on time than the biological species concept and has been criticized as being vague and not subject to observational test (O'Hara 1994).

The **phylogenetic species** has been defined as "the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent" (Cracraft 1983). Critics of this concept note that the definition of diagnosability is vague; if examined carefully, characters can be found to diagnose virtually any population, especially if molecular techniques are used. This would tend to greatly increase the number of species (O'Hara 1994).

Thus, depending on which species definition and assumptions a systematist uses, species are not necessarily equivalent. O'Hara (1994) noted that many species are easy to delimit, and

> in those cases where they are not, the difficulty that arises illustrates well the special *historical* character of the evolutionary process. . . . Because evolutionary history is something we are still in the midst of, it will not always be possible for us to determine which varieties—which distinctive populations in nature—are temporary and which are permanent, and so our counts of species across space and through time will always have some measure of ambiguity in them that we cannot escape.

12.10.2. How Many Genes Are Involved in Speciation?

The genetic basis of speciation is assumed to be due to changes in more than one gene, but the number is unknown in most cases (Harrison 1991, Hollocher 1998). Changes in more than two genes have been considered the minimum required for reproductive isolation. Changes in segments of the genome, such as inversions or translocations, also can result in reproductive isolation. Most often, reproductive isolation is considered to be determined polygenically. Few genetic studies have been conducted on the genetic basis of reproductive isolation, or temporal or habitat isolation, so few generalizations can be made about the number of genes responsible for speciation (Coyne 1992). A variety of characters can contribute to speciation, including hybrid sterility, hybrid inviability, interspecific mate discrimination, and interspecific divergence in secondary sexual characters (Hollocher 1998).

Speciation may occur rather rapidly under some circumstances. Higgie et al. (2000) showed that artificial selection can produce the kind of isolation that separates species in the wild, and that it can do so within nine generations under laboratory conditions. *Drosophila serrata* and *D. birchii* are sibling species in Australia that are very similar in morphology and can produce viable and fertile hybrid progeny in the laboratory. In the field, these two species can be found in the same area, but rarely interbreed. Where their geographic ranges overlap, the two differ in the mix of hydrocarbons found on their cuticle, which is important in mate choice. Populations of *D. serrata* found in regions of Australia where *D. birchii* do not occur have a different set of hydrocarbons. This suggests that selection to reduce mating between the species has occurred where the two populations overlap.

Under laboratory conditions, cuticular hydrocarbons of allopatric *D. serrata* populations collected from the field evolved within nine generations to resemble those of the sympatric populations when held with populations of *D. birchii*. This experiment in artificial sympatry indicates how rapidly mate recognition systems can change. However, the experiment does not indicate whether selection on mate recognition was a component of the actual speciation event resulting in *D. serrata* and *D. birchii*.

Drosophila melanogaster populations from Zimbabwe and populations from other continents have been shown to be reproductively isolated. In the presence of males of their own kind, females from Zimbabwe do not mate with males from elsewhere; reciprocal mating is reduced, as well. The genes for this behavior apparently are found on autosomes II and III (Wu et al. 1995). The data suggest that these populations are in the "early stages of speciation" and that it is "driven by sexual selection" (Wu et al. 1995). Because *D. melanogaster* is so well known genetically, analyzing speciation should be especially tractable (Buckley et al. 1997).

A "speciation gene," called *Odysseus*, was cloned, sequenced, and compared between two closely related *Drosophila* species (*simulans* and *mauritiana*) (Ting et al. 2000). Ting et al. (2000) were testing the hypothesis that genomes may contain ancient polymorphisms, or gene introgression could have occurred, so that molecular phylogenies may not reflect reproductive isolation accurately. Rather, "speciation genes" may be better indicators of phylogenetic history. *Odysseus* is the cause of hybrid male sterility in the *D. simulans* clade; Ting et al. (2000) compared *Odysseus* and microsatellite sequences from *simulans*, *mauritiana*, and *sechelia* (with *D. melanogaster* as an outgroup). The results indicated that the genome (as indicated by 39 microsatellite loci) can "indeed be a mosaic of regions of different genealogies among closely related species, because of shared ancient polymorphism and /or introgressions." The sequences of *Odysseus*, by contrast, provided a clear resolution of species because there were extensive amino acid differences. The authors were surprised to find very different "resolutions between the genealogical trees of regions of DNA less than 3 kb apart. The hitchhiking process, either in removing ancient polymorphisms or in excluding cointrogressions of tightly linked variations, must have been relatively ineffectual over a longer distance." The authors raise the possibility that "diverging species that remain incompletely isolated reproductively (such as *D. simulans* and *D. mauritiana*) may be permeable to introgression over a large portion of their genomes." Because only a small region near each "speciation locus" is impermeable, the exchange may continue for some time until reproductive isolation is complete.

Many questions about speciation remain unanswered (Howard and Berlocher 1999). How important are conventional gene mutations compared to novel genetic elements, such as repeated sequences, microorganisms such as *Wolbachia*, or transposable elements? How often are "speciation" genes altered in their coding sequence compared to changes in noncoding regulatory regions? How often is reproductive isolation based on polyploidy, or on chromosomal rearrangements of chromosomes? If transposable elements, polyploidy, or infectious microorganisms such as *Wolbachia* cause speciation, would they produce a rapid change without significant genetic change in their host arthropod? Molecular analyses may provide answers to some of these questions (Coyne 1992, Templeton 2001). Molecular mapping could determine how often reproductive isolation has a simple or complex genetic basis. Cloning and characterizing genes important in speciation may provide information on how reproductive isolation occurs at the molecular level.

Templeton (2001) proposes using phylogeographic analyses of gene trees to test species status and processes. A gene tree is a reconstruction of the history of the genetic variability

in a sample of homologous genes or DNA regions that have experienced little or no recombination. Templeton (2001) argues that gene trees can be used to understand the interface between intra- and interspecific evolution. He defines a "cohesion species" as an evolutionary lineage or set of lineages with genetic exchangeability and or ecological interchangeability. This species concept can be phrased in terms of null hypotheses that can be tested using gene trees.

12.10.3. Detecting Cryptic Species

One can argue that morphological methods are faster, easier, and cheaper than molecular methods for many taxonomic studies. However, molecular methods often provide the only method for detecting cryptic species. The ability to detect cryptic species may have ecological and economic importance.

RAPD-PCR may provide an inexpensive method for detecting cryptic species. For example, two populations of an encyrtid parasitoid, *Ageniaspis citricola*, were imported into the United States from Taiwan and Thailand as part of a classical biological control program directed against an invasive pest of citrus, *Phyllocnistis citrella* (Hoy et al. 2000). Slight differences in the biology and behavior of the two populations led us to evaluate them with RAPD-PCR, and the results indicated the two populations were genetically distinct (Hoy et al. 2000). Subsequently, analysis of two highly conserved *actin* genes confirmed the distinctiveness of these populations. Analysis of ribosomal ITS2 sequences also indicated the two populations are different (Alvarez and Hoy 2002). Interestingly, multiple clones of the ITS2 region were sequenced from individuals, and the intraindividual sequence variation observed was sometimes greater than sequence variation between individuals. This variability in sequence and length of the ITS2 region in the *Ageniaspis* populations suggests that concerted evolution has not homogenized all copies of the rRNA genes within these individuals and populations. Yet, despite this variability, the ITS2 region was informative phylogenetically (Alvarez and Hoy 2002).

Another economically important example of cryptic species involves a mite, *Varroa*, that is a parasite of honey bees (Oldroyd 1999, Anderson and Trueman 2000). *Apis mellifera* originally was restricted to Europe, while *A. cerana* was found in Asia. These sibling bee species came into contact after 1905 when the trans-Siberian railroad was completed. *Varroa* moved onto *A. mellifera* sometime in the past century and created serious problems for bee keepers in many locations around the world. *Apis cerana* is parasitized by *Varroa*, which suck the blood of developing bees and adults, but the effects of *Varroa* on *A. cerana* are relatively mild.

Until recently, it was believed that *Varroa* was a homogeneous species (called *jacobsoni*), but molecular studies (RAPD-PCR and mitochondrial DNA analyses) indicated that there is considerable hidden genetic variability within *Varroa* populations (Kraus and Hunt 1995, de Guzman et al. 1997, 1998). In fact, molecular data suggest there were multiple introductions of *Varroa* into the Western Hemisphere (de Guzman et al. 1999). Furthermore, molecular data suggest that *A. cerana* is, in fact, attacked by at least two cryptic species of *Varroa* (*jacobsoni* and *underwoodi*).

In the Western Hemisphere, two introduced "strains" of *Varroa* coexist. One is highly destructive to European bees (*A. mellifera*) and one relatively benign (Oldroyd 1999). Bee breeders are in the process of genetically selecting strains of *A. mellifera* that have a shorter development time, which could make them less vulnerable to attack by *Varroa*. Another strategy is to select bees for behavioral resistance to *Varroa*. Conducting such selections

in an effective manner requires that the "virulent" *Varroa* strains (or cryptic species) be used during the selection process. Thus, identifying cryptic *Varroa* species may explain differences in the effects of "*Varroa*" on bee populations and could be crucial to developing a practical solution(s) to the problem.

Some Relevant Journals

Biochemical Systematics and Ecology. Pergamon Press, Elmsford, NY.
Cladistics. Academic Press, San Diego.
Journal of Molecular Evolution. Springer-Verlag, New York.
Molecular Biology and Evolution. University of Chicago Press, Chicago.
Molecular Phylogenetics and Evolution. Academic Press, San Diego.

References Cited

Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prudhomme, and R. de Rosa (2000). The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. USA* **97**: 4453–4456.

Alvarez, J. M., and M. A. Hoy (2002). Evaluation of the ribosomal ITS2 DNA sequences in separating closely related populations of the parasitoid *Ageniaspis* (Hymenoptera: Encyrtidae). *Ann. Entomol. Soc. Am.* **95**: 250–256.

Anderson, D. L., and J. W. H. Trueman (2000). *Varroa jacobsoni* (Acari: Varroidae) is more than one species. *Exp. Appl. Acarol.* **24**: 165–189.

Attimonelli, M., N. Altamura, R. Benne, A. Brennicke, J. M. Cooper, D. D'Ellia, A. de Montalvo, et al. (2000). MitBASE: a comprehensive and integrated mitochondrial DNA database. The present status. *Mol. Biol. Evol.* **13**: 150–160.

Aubert, J., and M. Solignac (1990). Experimental evidence for mitochondrial DNA introgression between *Drosophila* species. *Evolution* **44**: 1272–1282.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* **18**: 489–522.

Ayala, F. J. (1986). On the virtues and pitfalls of the molecular evolutionary clock. *Heredity* **77**: 226–235.

Bachmann, L., J. M. Schibel, M. Raab, and D. Sperlich (1993). Satellite DNA as a taxonomic marker. *Biochem. System. Ecol.* **21**: 3–11.

Ballard, J. W. O. (2000). Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Mol. Evol.* **51**: 48–63.

Ballard, J. W. O., and M. Kreitman (1995). Is mitochondrial DNA a strictly neutral marker? *Trends Ecol. Evol.* **10**: 485–488.

Baxevanis, A. D., and B. F. F. Ouellette (2001). *Bioinformatics. A Practical Guide to the Analysis of Genes and Proteins*. 2nd ed., Wiley Interscience, New York.

Beard, C. B., D. M. Hamm, and F. S. Collins (1993). The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization and comparison with the mitochondrial sequences of other insects. *Insect Mol. Biol.* **2**: 103–124.

Boore, J. L., and W. M. Brown (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Gen. Dev.* **8**: 668–674.

Boore, J. L., D. Lavrov, and W. M. Brown (1998). Gene translocation links insects and crustaceans. *Nature* **392**: 667–668.

Brown, T. A., Ed. (1991). *Molecular Biology LabFax*. Bios Scientific Publ., Oxford, UK.

Brown, J. R., and W. F. Doolittle (1997). Archaea and the prokaryote-to-eukaryote transition. *Micro. Mol. Biol. Rev.* **61**: 456–502.

Buckley, S. H., T. Tregenza, and R. K. Butlin (1997). Speciation and signal trait genetics. *Trends Ecol. Evol.* **12**: 299–301.

Buckley, T. R., C. Simon, P. K. Flook, and B. Misof (2000). Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. *Insect Mol. Biol.* **9**: 565–580.

Burmester, T., H. C. Massey, Jr., S. O. Zakharkin, and H. Bene (1998). The evolution of hexamerins and the phylogeny of insects. *J. Mol. Evol.* **47**: 93–108.

Bushman, F. (2002). *Lateral DNA Transfer. Mechanisms and Consequences*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Caterino, M. S., S. Cho, and F. A. H. Sperling (2000). The current state of insect molecular systematics: a thriving Tower of Babel. *Annu. Rev. Entomol.* **45**: 1–54.

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol. Rev.* **73**: 2003–266.

Clary, D. O., and D. R. Wolstenholme (1985a). The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**: 252–271.

Clary, D. O., and D. R. Wolstenholme (1985b). The ribosomal RNA genes of *Drosophila* mitochondria DNA. *Nucleic Acids Res.* **13**: 4029–4045.

Cook, C. E., M. L. Smith, M. J. Telford, A. Bastianello, and M. Akam (2001). Hox genes and the phylogeny of the arthropods. *Curr. Biol.* **11**: 759–763.

Coyne, J. A. (1992). Genetics and speciation. *Nature* **355**: 511–515.

Cracraft, J. (1983). Species concepts and speciation analysis. *Curr. Ornithol.* **1**: 159–187.

Crease, T. J. (1999). The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* **233**: 89–99.

Cross, N. C. P., and G. A. Dover (1987). Tsetse fly rDNA: an analysis of structure and sequence. *Nucleic Acids Res.* **15**: 15–30.

Crozier, R. H., and Y. C. Crozier (1993). The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**: 97–117.

Dacks, J. B., and W. F. Doolittle (2001.) Reconstructing/deconstructing the earliest Eukaryotes: how comparative genomics can help. *Cell* **107**: 419–425.

Daly, H. V., J. T. Doyen, and A. H. Purcell III. (1998). *Introduction to Insect Biology and Diversity*. 2nd ed. Oxford Univ. Press, Oxford, UK.

de Guzman, L. I. , T. E. Rinderer, and J. A. Stelzer (1997). DNA evidence of the origin of *Varroa jacobsoni* Oudemans in the Americas. *Biochem. Genet.* **35**: 327–335.

de Guzman, L. I., T. E. Rinderer, J. A. Stelzer, and D. Anderson (1998). Congruence of RAPD and mitochondrial markers in assessing *Varroa jacobsoni* genotypes. *J. Apicult. Res.* **37**: 49–51.

de Guzman, L. I., T. E. Rinderer, and J. A. Stelzer (1999). Occurrence of two genotypes of *Varroa jacobsoni* Oud. in North America. *Apidologie* **30**(1): 31–36.

De Queiroz, K., and J. Gauthier (1994). Toward a phylogenetic system of biological nomenclature. *Trends Ecol. Evol.* **9**: 27–31.

Doolittle, R. F., Ed. (1990). *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Methods in Enzymology. Vol. 183. Academic Press, New York.

Doolittle, W. F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14**: 307–311.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284**: 2124–2128.

Doolittle, W. F. (2000). Uprooting the tree of life. *Sci. Am.*, February, 72–77.

Dotson, E. M., and C. B. Beard (2001). Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol. Biol.* **10**: 205–215.

Dowling, T. E., C. Moritz, J. D. Palmer, and L. H. Rieseberg (1996). Nucleic acids III: Analysis of fragments and restriction sites, pp. 249–320. In: *Molecular Systematics*. 2nd ed. D. M. Hillis, C. Moritz, and B. K. Mable, Eds. Sinauer Assoc., Sunderland, MA.

Eernisse, D. J. (1998). A brief guide to phylogenetic software. *Trends Genet.* **14**: 473–475.

Eickbush, T. H. (2002). R2 and related site-specific non-long terminal repeat retrotransposons, pp. 813–835. In: Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert, and A.M. Lambowitz, Eds. ASM Press, Washington, DC.

Ereshefsky, M. (2001). *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge Univ. Press, Cambridge, UK.

Feder, J. L., C. A. Chilcote, and G. L. Bush (1988). Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**: 61–64.

Feder, J. L., J. B. Roethele, B. Wlazlo, and S. H. Berlocher (1997). Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. *Proc. Natl. Acad. Sci. USA* **94**: 11417–11421.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**: 521–565.

Felsenstein, J. (2002). *Inferring Phylogenies*. Sinauer Assoc., Sunderland, MA.

Ferraris, J. D., and S. R. Palumbi, Eds. (1996). *Molecular Zoology, Advances, Strategies, and Protocols*. Wiley-Liss, New York.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace, and R. A. Raff (1988). Molecular phylogeny of the animal kingdom. *Science* **239**: 748–753.

Fitch, W. M., and F. J. Ayala (1994). The superoxide dismutase molecular clock revisited. *Proc. Natl. Acad. Sci. USA*: 91: 6802–6807.

Flook, P. K., C. H. Rowell, and G. Gellissen (1995). The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* **41**: 928–941.

Fortna, A., and K. Gardiner (2001). Genomic sequence analysis tools: a user's guide. *Trends Genet.* **17**: 158–164.

Fox, D. L., D. C. Fisher, and L. R. Leighton (1999). Reconstructing phylogeny with and without temporal data. *Science* **284**: 1816–1819.

Friedlander, T. P., J. C. Regier, and C. Mitter (1992). Nuclear gene sequences for higher level phylogenetic analysis: 14 promising candidates. *Syst. Biol.* **41**: 483–490.

Friedman, R., and A. L. Hughes (2001). Gene duplication and the structure of eukaryotic genomes. *Genome Res.* **11**: 373–381.

Galant, R., and S. B. Carroll (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* **415**: 910–913.

Gibson, G., and S. V. Muse (2002). *A Primer of Genome Science*. Sinauer Assoc., Sunderland, MA.

Golding, G. B., and R. S. Gupta (1995). Protein-based phylogenies support a chimeric origin for the eukaryotic genome. *Mol. Biol. Evol.* **12**: 1–6.

Goldstein, D. B., and P. H. Harvey (1999). Evolutionary inference from genomic data. *BioEssays* **21**: 148–156.

Graur, D., and W. H. Li (2000). *Fundamentals of Molecular Evolution*. 2nd ed. Sinauer Assoc., Sunderland, MA.

Gray, M. W. (1989). Origin and evolution of mitochondrial DNA. *Annu. Rev. Cell Biol.* **5**: 25-50.

Gray, M. W., G. Burger, and B. F. Lang (1999). Mitochondrial evolution. *Science* **283**: 1476–1481.

Green, E. D. (2001). Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* **2**: 573–583.

Gribskov, M., and J. Devereux, Eds. (1991). *Sequence Analysis Primer*. Stockton Press, New York.

Hall, B. G. (2001). *Phylogenetic Trees Made Easy. A How-To Manual for Molecular Biologists*. Sinauer Assoc., Sunderland, MA.

Hall, H. G. (1998). Restriction fragment length polymorphism (RFLP) analyses, pp. 333-366. In: Methods in Molecular Biology 50: *Species Diagnostics Protocols: PCR and other Nucleic Acid Methods*. J. P. Clapp, Ed., Humana, Totowa, NJ.

Hames, B. D., and D. Rickwood, Eds. (1981). *Gel Electrophoresis of Proteins. A Practical Approach*. IRL Press, Oxford, UK.

Harrison, R. G. (1991). Molecular changes at speciation. *Annu. Rev. Ecol. Syst.* **22**: 281–308.

Harvey, P. H., A. J. Leigh Brown, J. Maynard Smith, and S. Nee, Eds. (1996). *New Uses for New Phylogenies*, Oxford Univ. Press, Oxford, UK.

Hennig, W. (1966). *Phylogenetic Systematics*. Univ. Illinois Press, Urbana, IL.

Hickson, R. E., C. Simon, A. Cooper, G. S. Spicer, J. Sullivan, and D. Penny (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* **13**: 150–160.

Higgie, M., S. Chenoweth, and M. W. Blows (2000). Natural selection and the reinforcement of mate recognition. *Science* **290**: 519–520.

Hillis, D. M., and C. Moritz (1990). An overview of applications of molecular systematics, pp. 502–515. In: *Molecular Systematics*. D. M. Hillis, and C. Moritz, Eds. Sinauer Assoc., Sunderland, MA.

Hillis, D. M., A. Larson, S. K. Davis, and E. A. Zimmer (1990). Nucleic Acids III: sequencing, pp. 318–370. In: *Molecular Systematics*. D. M. Hillis, and C. Moritz, Eds. Sinauer Assoc., Sunderland, MA.

Hillis, D. M., C. Morita, and B. K. Mable, Eds. (1996). *Molecular Systematics*. 2nd ed. Sinauer Assoc., Sunderland, MA.

Holbrook, F. R., W. J. Tabachnick, E. T. Schmidtmann, C. N. McKinnon, R. J. Bobian, and W. C. Grogan (2000). Sympatry in the *Culicoides variipennis* complex (Diptera: Certapogonidae): A taxonomic reassessment. *J. Med. Entomol.* **37**: 65–76.

Hollocher, H. (1998). Reproductive isolation in *Drosophila*: how close are we to untangling the genetics of speciation? *Curr. Opin. Genet. Dev.* **8**: 709–714.

Howard, D. J., and S. T. Berlocher, Eds. (1999). *Endless Forms: Species and Speciation*. Oxford Univ. Press, Oxford, UK.

Howe, C. J., and E. S. Ward, Eds. (1989). *Nucleic Acids Sequencing. A Practical Approach*. IRL Press, New York.

Hoy, M. A., A. Jeyaprakash, R. Morakote, P. K. C. Lo, and R. Nguyen (2000). Genomic analyses of two populations of *Ageniaspis citricola* (Hymenoptera: Encyrtidae) suggest that a cryptic species may exist. *BioControl* **17**: 1–10.

Huelsenbeck, J. P., and B. Rannala (1997). Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**: 227–232.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**: 2310–2314.

Hwang, U. W., C. J. Park, T. S. Yong, and W. Kim (2001). One-step PCR amplification of complete arthropod mitochondrial genomes. *Mol. Phylogen. Evol.* **19**: 345–352.

Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White, Eds. (1990). *PCR Protocols, A Guide to Methods and Applications*. Academic Press, San Diego, CA.

Jain, R., M. C. Rivera, and J. A. Lake (1999). Horizontal transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96**: 3801–3806.

Katz, L. A. (1998). Changing perspectives on the origin of eukaryotes. *Trends Ecol. Genet.* **13**: 493–497.

Katz, L. A. (1999). The tangled web: gene genealogies and the origin of Eukaryotes. *Am. Natur. Suppl.* **154**: S137–S145.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**: 624–626.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press, Cambridge, UK.

Kimura, M. (1987). Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.* **26**: 24–33.

Klass, K. D., O. Zompro, N. P. Kristensen, and J. Adis (2002). Mantophasmatodea: a new insect order with extant members in the Afrotropics. Science Express. www.sciencexpress.org 18 April 2002,10.1126 Science1069397.

Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata, and S. I. Chigusa (1990). Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics* **126**: 657–663.

Kraus, B., and G. Hunt (1995). Differentiation of *Varroa jacobsoni* Oud. populations by random amplification of polymorphic DNA (RAPD). *Apidologie* **26**: 283–290.

Kreitman, M. (1996). The neutral theory is dead. Long live the neutral theory. *BioEssays* **18**: 678–683.

Kukalova-Peck, J. (1991). Fossil history and the evolution of hexapod structures, pp. 141–179. In: *The Insects of Australia*. Vol. I, 2nd ed. Melbourne Univ. Press, Melbourne, Australia.

Kumar, A., and K. S. Rai (1990). Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera: Culicidae). *Hereditas* **113**: 277–289.

Labandeira, C. C., and J. J. Sepkoski, Jr. (1993). Insect diversity in the fossil record. *Science* **261**: 310–315.

Labandeira, C. C., B. S. Beall, and F. M. Hueber (1988). Early insect diversification: evidence from a lower Devonian bristletail from Quebec. *Science* **242**: 913–916.

Lanave, C., F. :Licciulli, M. De Robertis, A. Marolla, and M. Attimonelli (2002). Update of AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Res.* **30**: 174–175.

Lang, B. F., M. W. Gray, and G. Burger (1999). Mitochondrial genome evolution and the origin of the Eukaryotes. *Annu. Rev. Genet.* **33**: 351–397.

Lansman, R. A., R. O. Shade, J. F. Shapiro, and J. C. Avise (1981). The use of restriction endonucleases to measure mitochondrial sequence relatedness in natural populations. III. Techniques and potential applications. *J. Mol. Evol.* **17**: 214–226.

Lansman, R. A., J. C. Avise, and M. D. Huettel (1983). Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **80**: 1969–1971.

Lavrov, D. V., J. L. Boore, and W. M. Brown (2000). The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol. Biol. Evol.* **17**: 813–824.

Lee, M. S. Y. (1999). Molecular phylogenies become functional. *Trends Ecol. Evol.* **14**: 177–178.

Lessinger, A. C., A. C. M. Junqueria, T. A. Lemos, E. L. Kemper, F. R. Da Silva, A. L. Vettore, P. Arruda, and A. M. L. Azeredo-Espin (2000). The mitochondrial genome of the primary screwworm fly *Cochliomyia homivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* **9**: 521–529.

Levine, M. (2002). How insects lose their limbs. *Nature* **415**: 848–849.

Lewis, D. L., C. L. Farr, and L. S. Kaguni (1995). *Drosophila melanogaster* mitochondrial DNA, completion of the nucleotide sequence and evolutionary comparisons. *Insect Mol. Biol.* **4**: 263–278.

Lewontin, R. C., and J. Hubby (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amounts of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**: 595–609.

Li, W. H., and D. Graur (1991). *Fundamentals of Molecular Evolution*. Sinauer Publ., Sunderland, MA.

Liu, H., and A. T. Beckenbach (1992). Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylog. Evol.* **1**: 41–52.

Lunt, D. H., L. E. Whipple, and B. C. Hyman (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol. Ecol.* **7**: 1441–1455.

Maley, L. E., and C. R. Marshall (1998). The coming of age of molecular systematics. *Science* **279**: 505–506.

Manuel, M., M. Kruse, W. E. G. Muller, and Y. Le Parco (2000). The comparison of β-thymosin homologues among Metazoa supports and arthropod-nematode clade. *J. Mol. Evol.* **51**: 378-381.

Margulis, L. (1970). *Origin of Eukaryotic Cells*. Yale Univ. Press, New Haven, CT.

Matsuura, E. T., H. Fukuda, and S. I. Chigusa (1991). Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Reunion. *Genet. Res., Camb.* **57**: 123–126.

Maxon, L. R., and R. D. Maxon (1990). Proteins II: Immunological techniques, pp. 127–155. In: *Molecular Systematics*. D. M. Hillis, and C. Moritz, Eds. Sinauer Assoc., Sunderland, MA.

May, B. (1992). Starch gel electrophoresis of allozymes, pp. 1–27. In: *Molecular Genetic Analysis of Populations. A Practical Approach*. A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

Mayr, E. (1970). *Populations, Species, and Evolution*. Harvard Univ. Press, Cambridge, MA.

Mayr, E. (1998). Two empires or three? *Proc. Natl. Acad. Sci. USA* **95**: 9720–9723.

Mayr, E., and P. D. Ashlock (1991). *Principles of Systematic Zoology*. 2nd ed. McGraw-Hill, New York.

Mitchell, S. E., A. F. Cockburn, and J. A. Seawright (1993). The mitochondrial genome of *Anopheles quadrimaculatus*, species A: complete nucleotide sequence and gene organization. *Genome* **36**: 1058–1073.

Mooers, A. O., and E. C. Holmes (2000). The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* **15**: 365–369.

Moran, N. A., M. A. Munson, P. Baumann, and H. Ishikawa (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc. R. Soc. Lond. B* **253**: 167–171.

Moritz, C., and D. M. Hillis (1990). Molecular systematics: context and controversies, pp. 1–12. In: *Molecular Systematics*, D. M. Hillis, and C. Moritz, Eds., Sinauer Assoc., Sunderland, MA.

Moritz, C., T. E. Dowling, and W. M. Brown (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Evol. Syst.* **18**: 269–92.

Mount, D. W. (2001). *Bioinformatics. Sequence and Genome Analysis*. Cold Spring Harbor Lab. Press, NY.

Murphy, R. W. J. W. Sites, D. G. Buth, and C. Haufler (1996.) Proteins: Isozyme electrophoresis, pp. 51–120. In: *Molecular Systematics*. 2nd ed. D. M. Hillis, C. Moritz, and B. K. Mable, Eds. Sinauer Assoc., Sunderland, MA.

Nardi, F., A. Carapelli, P. P. Franciulli, R. Dallai, and F. Frati (2001). The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. *Mol. Biol. Evol.* **18**: 1293–1304.

Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu. Rev. Genet.* **30**: 371–403.

O'Hara, R. J. (1994). Evolutionary history and the species problem. *Am. Zool.* **34**: 12–22.

Ohno, S. (1970). *Evolution by Gene Duplication*. Springer, New York.

Ohta, T. (1996). The current significance and standing of neutral and nearly neutral theories. *BioEssays* **18**: 673–677.

Ohta, T. (2000a). Evolution of gene families. *Gene* **259**: 45–52.

Ohta, T. (2000b). Mechanisms of molecular evolution. *Phil. Trans. R. Soc. Lond. B* **355**: 1623–1626.

Oldroyd, B. P. (1999). Coevolution while you wait: *Varroa jacobsoni*, a new parasite of western honeybees. *Trends Ecol. Evol.* **14**: 312–315.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.

Palumbi, S. R. (1996). Nucleic acids II: The polymerase chain reaction, pp. 205–248, In: D. M. Hillis, C. Moritz, and B. K. Mable, Eds., *Molecular Systematics*. 2nd ed. Sinauer Assoc., Sunderland, MA.

Pashley, D. P., and L. D. Ke (1992). Sequence evolution in mitochondrial ribosomal and ND-1 genes in Lepidoptera: implications for phylogenetic analyses. *Mol. Biol. Evol.* **9**: 1061–1075.

Pasteur, N., G. Pasteur, F. Bonhomme, J. Catalan, and J. Britton-Davidian (1988). *Practical Isozyme Genetics*. Ellis Horwood, Chichester, UK.

Pennisi, E. (2001). Linnaeus's last stand? *Science* **291**: 2304–2307.

Philippe, H., and P. Forterre (1999). The rooting of the universal tree of life is not reliable. *J. Mol. Evol.* **49**: 509–523.

Philippe, H., A. Germot, and D. Moreira (2000). The new phylogeny of eukaryotes. *Curr. Opin. Genet. Dev.* **10**: 596–601.

Post, R. J., P. K. Flook, and M. D. Wilson (1992). DNA analysis in relation to insect taxonomy, evolution, and identification, pp. 21–34. In: *Insect Molecular Science*. J. M. Crampton, and P. Eggleston, Eds. Academic Press, London, UK.

Powell, J. R. (1991). Monophyly/paraphyly/polyphyly and gene/species trees: an example from *Drosophila. Mol. Biol. Evol.* **8**: 892–896.

Powell, J. R., and A. Caccone (1990). The TEACL method of DNA–DNA hybridization: technical considerations. *J. Mol. Evol.* **30**:267–272.

Reeck, G. R., C. deHaen, D. C. Teller, R. F. Doolittle, W. M. Fitch, R. E. Dickerson, P. Chambon, A. D. McLachlan, E. Margoliash, T. H. Jukes, and E. Zuckerkandl (1987). "Homology" in proteins and nucleic acids: a terminology muddle and a way out of it. *Cell* **50**: 667.

Regier, J. C., and J. W. Shultz (1997). Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. *Mol. Biol. Evol.* **14**: 902–913.

Rice, W. R., and E. E. Hostert (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution* **47**: 1637–1653.

Rivera, M. C., R. Jain, J. E. Moore, and J. A. Lake (1998). Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* **95**: 6239–6244.

Rodriguez-Trelles, F., R. Tarrio, and F. J. Ayala (2001). Erratic over dispersion of three molecular clocks: GPDH, SOD, and XDH. *Proc. Natl. Acad. Sci. USA* **98**: 11405–11410.

Roehrdanz, R. L., and M. E. DeGrugillier (1998). Long sections of mitochondrial DNA amplified from fourteen orders of insects using conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **91**: 771–778.

Roger, A. J. (1999). Reconstructing early events in eukaryotic evolution. *Am. Natur.* **154**: S146–S163.

Rokas, A., and P. W. H. Holland (2000). Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15**: 454–459.

Ronshaugen, M., N. McGinnis, and W. McGinnis (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* **415**: 914–917.

Sankoff, D. (2001). Gene and genome duplication. *Curr. Opin. Genet. Dev.* **11**: 681–684.

Satta, Y., and N. Takahata (1990). Evolution of *Drosophila* mitochondrial DNA and the history of the *melanogaster* subgroup. *Proc. Natl. Acad. Sci. USA* **87**: 9558–9562.

Saville, B. J., Y. Kohli, and J. B. Anderson (1998). mtDNA recombination in a natural population. *Proc. Natl. Acad. Sci. USA* **95**: 1331–1335.

Sessions, S. K. (1996). Chromosomes: molecular cytogenetics, pp. 121–168. In: D. M. Hillis, C. Moritz, and B. K. Mable, Eds., *Molecular Systematics*. 2nd ed. Sinauer Assoc., Sunderland, MA.

Shao, R., N. J. H. Campbell, and S. C. Barker (2001). Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Pthiraptera). *Mol. Bio. Evol.* **18**: 858–865.

Shoemaker, J. S., I. S. Painter, and B. S. Weir (1999). Bayesian statistics in genetics. A guide for the uninitiated. *Trends Genet.* **15**: 354–358.

Simon, C. (1991). Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA, pp. 35–71. In: *Molecular Techniques in Taxonomy*. G. M. Hewitt, A. W. B. Johnston, and J. P. W. Young, Eds. Springer-Verlag, Berlin, Germany.

Simon, C., S. Paabo, T. D. Kocher, and A. C. Wilson (1990). Evolution of mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction, pp. 235–244. In: *Molecular Evolution*. M. T. Clegg, and S. J. O'Brien, Eds. UCLA Symp. Mol. Cell. Biol. New Series, Vol. 122. A. R. Liss Publ., New York.

Simon, C., A. Franke, and A. Martin (1991). The polymerase chain reaction: DNA extraction and amplification, pp. 329–355. In: *Molecular Techniques in Taxonomy*. G. M. Hewitt, W. Johnston, and J. P. W. Young, Eds. Springer-Verlag, Berlin, Germany.

Simpson, G. G. (1961). *Principles of Animal Taxonomy*. Columbia Univ. Press, New York.

Spanos, L., G. Koutroumbas, M. Kotsyfakis, and C. Louis (2000). The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Mol. Biol.* **9**: 139–144.

Steel, M., and D. Penny (2000). Parsimony, likelihood and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* **17**: 839–850.

Stewart, C. B. (1993). The powers and pitfalls of parsimony. *Nature* **361**: 603–607.

Swofford, D. L., and G. J. Olsen (1990). Phylogeny reconstruction, pp. 411–501. In: *Molecular Systematics*. D. M. Hillis, and C. Moritz, Eds. Sinauer Assoc., Sunderland, MA.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis (1996). Phylogenetic inference, pp. 407–514. In: *Molecular Systematics*. 2nd ed. D. M. Hillis, C. Moritz, and B. K. Mable, Eds. Sinauer Assoc., Sunderland, MA.

Tabachnick, W. J. (1992). Genetic differentiation among populations of *Culicoides variipennis* (Diptera: Ceratopogonidae), the North American vector of bluetongue virus. *Ann. Entomol. Soc. Am.* **85**: 140–147.

Tabachnick, W. J. (1996). The genetics of *Culicoides variipennis* and the epidemiology of bluetongue disease in North America. *Annu. Rev. Entomol.* **45**: 20–40.

Tamura, K. (1992). The rate and pattern of nucleotide substitution in *Drosophila* mitochondrial DNA. *Mol. Biol. Evol.* **9**: 814-825.

Taylor, M. F. J., S. W. McKechnie, N. Pierce, and M. Kreitman (1993). The lepidopteran mitochondrial control region: structure and evolution. *Mol. Biol. Evol.* **10**: 1259–1272.

Templeton, A. R. (2001). Using phylogeographic analyses of gene trees to test species status and processes. *Mol. Ecol.* **10**: 779–791.

Ting, C. T., S. C. Tsaur, and C. I. Wu (2000). The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. USA* **97**: 5313–5316.

Turbeville, J. M., D. M. Pfeifer, K. G. Field, and R. A. Raff (1991). The phylogenetic status of arthropods, as inferred from 18S rRNA sequences. *Mol. Biol. Evol.* **8**: 669–686.

Van de Peer, Y., J. M. Neefs, P. De Rijk, and R. De Wachter (1993). Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. *J. Mol. Evol.* **37**: 221–232.

Wagner, A. (1998). The fate of duplicated genes: loss or new function? *BioEssays* **20**: 785–788.

Weir, B. S. (1990). *Genetic Data Analysis*. Sinauer Assoc., Sunderland, MA.

Werman, S. D., M. S. Springer, and R. J. Briteen (1996). Nucleic acids I; DNA–DNA hybridization, pp. 169–204. In: *Molecular Systematics*. 2nd ed. D. M. Hillis, C. Moritz, and B. K. Mable, Eds. Sinauer Assoc., Sunderland, MA.

Wheeler, W. C., M. Whiting, Q. D. Wheeler, and J. M. Carpenter (2001). The phylogeny of the extant hexapod orders. *Cladistics* **17**: 113–169.

Whelan, S., P. Lio, and N. Goldman (2001). Molecular phylogenetics: state-of-the art methods for looking into the past. *Trends Genet.* **17**: 262–272.

White, M. J. D. (1973). *Animal Cytology and Evolution*. 3rd ed. Cambridge Univ. Press, Cambridge, UK.

White, M. J. D. (1978). *Modes of Speciation*. W. H. Freeman, San Francisco.

White, P. S., and L. D. Densmore III. (1992). Mitochondrial DNA isolation, pp. 29–58. In: *Molecular Genetic Analysis of Populations. A Practical Approach*. A. R. Hoelzel, Ed. IRL Publ., Oxford, UK.

Wilson, A. C., H. Ochman, and E. M. Prager (1987). Molecular time scale for evolution. *Trends Genet.* **3**: 241–247.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**: 227–271.

Woese, C. R. (2000). Interpreting the universal phylogenetic tree. Proc. *Natl. Acad. Sci. USA* **97**: 8392–8396.

Woese, C. R., O. Kandler, and M. L. Wheelis (1990). Towards a natural system of organisms: proposal for domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**: 4576–4579.

Wu, C. I., H. Hollocher, D. J. Begun, C. F. Aquadro, Y. Xu, and M. L. Wu (1995). Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. *Proc. Natl. Acad. Sci. USA* **92**: 2519–2523.

Yang, Z., and J. P. Bielawski (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**: 496–503.

Zuckerkandl, E., and L. Pauling (1965). Evolutionary divergence and convergence in proteins, pp. 97–166. In: *Evolving Genes and Proteins*. V. Bryson, and H. J. Vogel, Eds. Academic Press, New York.

Some Relevant Web Sites

Bioinformatics Education, a collaboration between universities in Sweden, Singapore, Australia, Sweden, South Africa and the United States, provide bioinformatics training: http://s-star.org

GenBank: www.ncbi.nim.hih.gov
The MrBayes program: http://brahms.biology.rochester.edu/software.html
Nature Genome Gateway: www.nature.com/nature/

PhyloCode Web site: www.ohio.edu/phylocode

TreeBASE, a database of the data behind phylogenetic trees that have been published: http//www.treebase.org/treebase/index.html

Tree of Life, a tree containing all known organisms in one phylogenetic tree; also contains programs: http://phylogeny.arizona.edu/tree/programs

Programs for phylogenetics and population genetics: http://evolve.zps.ox.ac.uk/Home.html

13

Insect Population Ecology and Molecular Genetics

13.1. Overview

Molecular genetic techniques provide powerful tools for the study of insect biology, ecology, and population genetics in both natural and laboratory populations. Analysis of proteins, nuclear or mitochondrial DNA, and messenger RNA can be used to answer ecological questions. Analysis of proteins by electrophoresis has been useful with many insects, but some taxa with low levels of detectable genetic variation cannot be studied unless more sensitive DNA markers are used. DNA analyses can identify biotypes or sibling species, determine paternity, resolve whether hybridization or introgression occurs, and provide information on founder effects, population genetic structure, gene flow, inbreeding, genetic bottlenecks, dispersal, and selection intensity. The PCR is often used to reduce costs and allow large numbers of specimens to be sampled. Large amounts of genetic variation can be sampled rapidly and inexpensively in large numbers of individuals by the RAPD method of the PCR or by restriction enzyme digests of DNA amplified by the PCR (RFLP-PCR) or by the AFLP-PCR method. Although technically more challenging and expensive, techniques such as DNA fingerprinting using microsatellite DNA, heteroduplex analysis, or double-strand conformation polymorphism (DSCP) provide information on genetic variation at the individual and population levels. Recently, the use of DNA microarray (DNA chip) analysis has allowed researchers to evaluate the responses of plants to insect attack; microarray analysis shows promise of answering other important ecological questions.

In addition to improved molecular methods, improvements have been made in the statistical methods and population genetics models used to analyze data. The continued improvement in molecular techniques and analysis methods in population biology and ecology will provide opportunities to resolve both fundamental and applied questions in insect population ecology, population genetics, and pest management. The field has advanced to the point where we are no longer asking, "Can we do that?" It is now relevant to question in any molecular ecology project the "so what?" issue (Curtis 2002). Do the molecular methods solve real problems that are "not already solvable by simpler and cheaper methods" (Curtis 2002)?

13.2. Introduction

The fields of ecology and population genetics generally employ synthetic, rather than reductionist, research approaches. It is noteworthy that a reductionist approach to biology (molecular genetics) is providing a new tool for resolving problems in population genetics and ecology.

Insect ecology is the study of insects in their environment. Insect ecologists are concerned with the biology of groups of insects and with the pattern of relationships between

insects and their environment. Ecology thus is concerned with organisms, populations, and communities in ecosystems. Insect ecology is an important component of applied pest management programs and conservation, as well as of more fundamental value in elucidating ecological principles. Ecology sometimes is divided into autecology and synecology.

Autecology deals with the study of the individual or an individual species, its life history, its behavior, and its adaptations to the environment. **Synecology** investigates groups of organisms associated together as a unit (Odum 1971, Southwood 1978, Price 1997, Huffaker and Gutierrez 1999). At present, most ecological research employing molecular genetic techniques is autecological.

13.3. What Is Molecular Ecology?

Molecular ecology was defined by Paul Weiss (1950) to mean "the entire continuum of biological interactions between the molecular, cellular, organismal levels to the environment" (Lambert 1995). Molecular tools provide ecologists with diverse methods for evaluating these interactions and allow ecologists to answer questions that have been difficult to resolve using traditional methods.

Population geneticists study how genetic principles apply to entire populations (Hartl 1981, Real 1994, Hartl and Clark 1997). One of the most striking features of insect populations is their phenotypic diversity. An underlying assumption that population geneticists make is that this phenotypic diversity is matched by genetic diversity, and they attempt to deal with the phenotypic and genotypic differences among individuals.

Population genetics and population ecology have been distinct disciplines, but they have become less distinct in the past few years (Slatkin 1987, Kellenberger 1994, Mitton 1994, Real 1994, Hoffman et al. 1995, Loxdale and Lushai 1998, Sunnucks 2000, Black et al. 2001). The molecular analysis of genes and genetic systems may provide insights for both autecological and synecological studies because an insect's heredity determines its behavior and ability to survive in specific environments and communities. Changes in genes and gene frequencies in populations over evolutionary time are important for understanding both speciation and community structure (Hoffman et al. 1995).

The application of molecular genetic techniques to the study of insect population ecology will play an ever more significant role as insect ecologists discover the power, and limitations, of these new tools (Sunnucks 2000). Analyses could provide better understanding of biodiversity, biosafety issues, biotype ecology and evolution, colonization processes, conservation biology, diet analysis, dispersal, gene flow, geographical distribution, host–parasite interactions, hybridization or introgression, insect–plant interactions, kinship, paternity, pesticide resistance, population structure, species identity, sperm precedence, and vector biology (Mitton 1994, Hoffmann et al. 1995, Tabachnick and Black 1995, Schwartz et al. 1998, Rieseberg 1999, Bohonak 1999, Roderick 1996, Cavalli-Sforza 1998, Howard and Berlocher 1998, Davies et al. 1999b, Wang and Caballero 1999, Berticat et al. 2000, ffrench-Constant et al. 2000, Sunnucks 2000, Baldwin et al. 2001, Black et al. 2001, Hewitt 2001).

This chapter provides a survey of methods and several examples in which molecular methods have been used to answer different kinds of population biology and ecology problems.

13.4. Collecting Arthropods in the Field for Analysis

The ability to collect insects from the field may be regulated by state, federal or international regulations (Dick et al. 1993). Permits are required to collect organisms on certain federal lands (wildlife refuges, national parks and national monuments) and many state lands. Endangered and threatened species in the United States are regulated by the Fish and Wildlife Service of the Department of the Interior, and permits must be obtained to collect, possess, or transport any species on the "List of Endangered and Threatened Wildlife." Furthermore, the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) requires collecting permits and restricts collection or importation of any species on the international list.

Most countries limit importation of live organisms that could be harmful to crops, livestock, or humans. In the United States, the Plant Protection and Quarantine branch of the U.S. Department of Agriculture, Animal and Plant Health Inspection Service must be contacted prior to the importation and transportation of live plant pests, plant pathogens, and vectors of plant and animal disease. The Office of Biosafety of the Centers for Disease Control regulates the importation of agents of human disease or vectors that could harbor these agents.

13.5. Molecular Ecological Methods

Proteins, nuclear DNA, mitochondrial DNA, and DNA from symbionts such as *Wolbachia* can be analyzed to resolve ecological problems; each has advantages and disadvantages. The ease of study, amount of variation that can be detected, and cost differ with each target. Both single-copy and multiple-copy DNA sequences can be analyzed in nuclear DNA, which allows analysis of genetic variation at the individual, population, or species level. Mitochondrial (mt) DNA analysis provides sufficient variation for studies of individuals, populations, or species, depending upon the region of the mitochondrion analyzed.

Potential techniques include allozyme electrophoresis, restriction fragment length polymorphisms (RFLPs) of mtDNA or nuclear DNA, DNA fingerprinting by analysis of microsatellites, RAPD-PCR, heteroduplex analysis (HDA), amplified fragment length polymorphism (AFLP-PCR), sequencing of both nuclear and mt DNA, allele-specific PCR using standard or Long PCR protocols, and microarray analysis. These methods were described in Chapters 6, 7, and 8. Each varies in the time required, ease of execution, cost, and level of genetic variability that can be detected (Table 13.1).

DNA-based methods provide a way to examine DNA directly. Which DNA-based technique should be employed in a particular project depends on the goals and the level of DNA variation in the species of interest (Moritz et al. 1987, Mitton 1994, Roderick 1996, Loxdale and Lushai 1998). When attempting to resolve problems involving species not analyzed previously by that specific molecular method, it is difficult to predict whether a particular target DNA sequence or technique will be useful. This is because insect groups differ in the amount and type of genetic variability they contain. Thus, molecular ecology remains a developing field that is still refining the molecular tools employed and the statistical methods utilized for data analysis.

13.5.1. Allele-Specific PCR

Allele-specific PCR is rapid, easy, and appropriate for many population biology or ecology studies (Arnheim et al. 1990, see Chapter 8). Allele-specific PCR requires DNA sequence

Table 13.1. Molecular Methods That Can Be Used to Evaluate Insects in Ecological Studies

Technique	Level of discrimination • Type of data obtained (gene frequencies or base-pair changes)	+Advantages −Disadvantages	Selected references
AFLP-PCR	Detect differences in individuals and populations	+More reliable than RAPD-PCR, More user-friendly than RFLPs and microsatellites. Samples large amounts of genome. Sequence information not needed −May require relatively large amounts of clean DNA. Requires multiple operations	Mueller and Wolfenbarger 1999, see Chapter 8
Allele-specific PCR	Detect single-nucleotide differences in individuals and populations • Gene frequency and base-pair changes	+Small amounts of DNA required. Relatively rapid and inexpensive. Results can be visualized by staining with ethidium bromide/other labels −DNA sequence information needed for primers, or consensus primers. Each reaction provides information for only one locus	See Chapter 8; Arnheim et al. 1990, Innis et al. 1990, Erlich 1989
Long PCR variant	• Same as above	+Detects microbial DNA mixed with insect or plant DNA. Is 6 to 8 orders of magnitude more sensitive than standard PCR −More expensive because it uses two polymerases, including one that corrects errors. Care must be taken to avoid contamination because of increased sensitivity	Jeyaprakash and Hoy 2000, Hoy et al. 2001
DSCP; double-strand conformation polymorphism	• Detect changes in mobility of double-stranded DNA molecules on polyacryamide gels	+Can use PCR products for analysis. Rapid and inexpensive. Can identify new haplotypes for additional analysis −Some mutations don't produce changes in mobility, thus won't work with all PCR products. Sequence differences can't be estimated. Sequencing may be required	Hagerman 1990, Saad et al. 1994, Atkinson and Adams 1997
Microsatellites	Detect differences in individuals and populations in tandemly repeated units in nuclear DNA	+High levels of variation present in most insects	Bruford et al. 1992, Kirby 1990, Zane et al. 2002

continues

continued

Technique	Level of discrimination • Type of data obtained (gene frequencies or base-pair changes)	+Advantages −Disadvantages	Selected references
	• Neither gene frequency nor base-pair changes	−Comigrating bands may not be identical alleles at a locus. Relatively large amounts of clean, high molecular weight DNA required. Labeled probes required. Relatively expensive and labor intensive. Time and effort are required to identify repeated units	
RAPD-PCR	Differences in single nucleotides in nuclear DNA	+Useful for species with limited genetic information. Efficient. Relatively inexpensive. Requires little DNA	Hadrys et al. 1992, Haymer 1994
	• Gene frequency data	−Sensitive to DNA concentration. No genetic information on PCR products. Can yield nonreproducible products. Markers are dominant and heterozygotes may be difficult to identify. Incorrect scoring can occur if two different fragments comigrate	Edwards and Hoy 1993, MacPherson et al. 1993, Landry et al. 1993
RFLPs	Differences in single nucleotides detected by sequences recognized by restriction endonucleases in nuclear and mtDNA	+mt DNA most often analyzed. Standard probes are available	Aquadro et al. 1992, Dowling et al. 1990, Tegelstrom 1992
	• Gene frequency and changes in base pairs	−Requires large amounts of DNA. Usually requires radiolabeled probes. Single locus or several loci only analyzed. Relatively expensive and technically demanding	White and Densmore 1992
PCR-RFLPs	Differences in single nucleotide sequences in nuclear and mtDNA recognized by the specific restriction enzyme used	+Requires only a small amount of DNA. Can be visualized with EtBr. Less expensive and more sensitive than standard RFLPs	Karl and Avise 1993
	• Gene frequency data	−Specific primers required. Two separate procedures are required, making it more time consuming and expensive than allele-specific PCR	

continues

Table 13.1. Continued

Technique	Level of discrimination • Type of data obtained (gene frequencies or base-pair changes)	+Advantages −Disadvantages	Selected references
Protein electrophoresis	Detect changes in charged amino acids	+Inexpensive. Many protocols available. Produces codominant Mendelian characters of enzymes important in physiology	May 1992, Pasteur et al. 1988, Murphy et al. 1990
	• Gene frequency data	−Less sensitive than DNA tests. Number of tests that can be performed may be limited in small insects. Proteins subject to environmental influences	
Sequencing PCR-amplified DNA	Differences in single nucleotides of nuclear and mtDNA including coding and noncoding regions	+Relatively small amounts of DNA needed. High resolution possible. Some universal PCR primers available	Hoelzel and Green 1992
	• Gene frequency and changes in base pairs	−Time consuming and expensive. Relatively small portion of genome can be sampled. Technically more demanding than other methods. Not often used when large numbers of insects must be screened due to cost	

data so that primers can be developed (see Chapter 8), although some primers for ribosomal and mtDNA can be used on many species (Table 13.1).

Ecologists have long wanted to know who is eating whom in the environment. Allele-specific PCR is being investigated as a new tool to evaluate the diets of predators (Agusti et al. 1999, Zaidi et al. 1999, Hoogendorn and Heimpel 2001). Persistence time of prey DNA in the gut, size of target DNA sequences, and the abundance (single-copy versus multiple-copy genes) of prey DNA in predator guts vary by species and by temperature (which affects the digestion rate). This new application of allele-specific PCR will have to be developed for each predator–prey system, making validation of gut analyses time consuming.

A variation of allele-specific PCR, Long PCR, was described in Chapter 8. Because two DNA polymerases are used (one of which has the ability to proofread and correct errors in incorporation), Long PCR allows microbial DNA to be detected even when it is mixed with insect or plant DNA. Long PCR is especially useful for detecting low titers of *Wolbachia* DNA or of DNA from plant pathogens within insects (Jeyaprakash and Hoy 2000, Hoy et al. 2001).

13.5.2. Allozymes (Protein Electrophoresis)

Allozymes have been used to analyze mating systems (random versus assortative mating), inbreeding, genetic drift, hybridization, effective population size, degree of genetic differentiation among populations, and migration. Extensive protein (allozyme) variation

has been found in some natural insect populations. Exceptions often include haplo-diploid Hymenoptera and clonal organisms such as aphids (Crozier 1977, Lester and Selander 1979). Even for other insects, however, allozyme studies may underestimate the amount of variation, detecting only about 30% of the actual genetic diversity as determined by DNA-based methods.

Protein electrophoresis is one of the most cost-effective techniques available and is relatively easy to perform. For example, allozyme variability was used to identify Japan as the likely origin of the mosquito *Aedes albopictus* that recently colonized the United States and Brazil (Kambhampati et al. 1991). Allozymes also were used to demonstrate genetic differentiation between sympatrically occurring hawthorn and apple populations of the apple maggot fly *Rhagoletis pomonella* (Feder et al. 1988). Unfortunately, protein electrophoresis may not detect sufficient variation to answer some questions, and the number of analyses that can be performed with very small insects may be limited because of inadequate amounts of proteins (Table 13.1). Proteins are less stable than DNA and thus may be more sensitive to handling and storage problems.

13.5.3. Amplified Fragment Length Polymorphisms (AFLP-PCR)

AFLP is a PCR-based method to develop large numbers of markers for population analyses (Mueller and Wolfenbarger 1999; see also Chapter 8). AFLP-PCR is a relatively inexpensive and reliable method of identifying hundreds of genetic markers without requiring sequence information to develop primers. The main disadvantage is the difficulty in identifying homologous markers, which makes it less useful for studies in which it is important to identify heterozygous individuals. As with RAPD-PCR and microsatellite analyses, AFLP-PCR screens many different regions of the genome. AFLP markers have been useful for assessing genetic differences among individuals, populations, and species (Mueller and Wolfenbarger 1999). AFLP markers are thought to be more easily replicated than RAPD-PCR, although AFLP-PCR is more difficult to use and develop (Table 13.1).

13.5.4. Double-Strand Conformation Polymorphism (DSCP)

DSCP is used to detect single base changes in DNAs. DSCP is detected by differences in electrophoretic mobility in nondenaturing acrylamide gels of double-stranded DNA (Hagerman 1990, Saad et al. 1994, Atkinson and Adams 1997). Single base changes in the DNA may alter the curvature of the helical axis of double-stranded DNA, which could lead to changes in electrophoretic mobility. Not all mutations affect DNA curving, so some are undetected by this approach (Table 13.1).

DSCP markers could provide markers for species determination, kinship, and paternity analysis, as well as other aspects of population genetics that require variation from a rapidly evolving region of DNA. For example, Atkinson and Adams (1997) analyzed the mitochondrial control region of the termite *Nasutitermes corniger* by DSCP using PCR products and discovered highly variable markers for population studies. Higher levels of polymorphism were found by DSCP than by RFLP analysis. The DSCP data suggested that some termite colonies contained unrelated queens, each of which produced workers.

13.5.5. Heteroduplex Analysis (HDA)

Heteroduplex analysis combines some of the advantages of allele-specific PCR and RFLP methods with the advantage of direct sequence analysis to detect new alleles (Tang and

Unnasch 1997). This method originally was used to identify virus isolates or detect immune genotypes in humans. HDA detects changes in mobility on an electrophoresis gel of heteroduplex products formed between the strands of a probe DNA and a test DNA molecule. The number and type of mismatched bases within a given HDA determines the conformation and mobility of the DNA duplex during electrophoresis. HDA is sufficiently sensitive to detect single base changes in fragments up to 500 bp long.

HDA involves obtaining PCR products from the probe DNA and test DNA. The DNAs are mixed, then denatured by heating; heteroduplex and homoduplex products are formed during the cooling of the sample to room temperature. This results in four products: the probe and sample DNA because the probe and sample strands reanneal to themselves (homoduplexes); and two heteroduplex products, each comprising one strand from the probe and one from the test DNA. These homo- and heteroduplex DNAs are separated by polyacrylamide gel electrophoresis, and the separated products are detected by ethidium bromide staining. The relative amount of retardation of the heteroduplex products compared to the homoduplexes reflects the number and type of mismatched nucleotides between the probe and test DNA. This allows new alleles to be detected, which can be further analyzed by sequencing to determine their relationship to previously identified sequences. HDA is sufficiently rapid that multiple individuals can be screened, allowing one to determine the allele frequency in the population.

Tang and Unnasch (1997) suggested HDA be applied to understanding medically important vectors of disease. They argue that HDA is simple, rapid, inexpensive, and capable of detecting small differences among DNA sequences. Disadvantages to HDA include the difficulty of measuring differences in the mobility of the sequences on the gels; such differences are only a rough estimate of the relative genetic distance between two sequences. Furthermore, some changes may yield a bend in the DNA molecule, resulting in a disproportionate change in gel mobility. Further analysis of this method will resolve its value for molecular ecology.

13.5.6. Microarrays

DNA microarrays (also known as gene chips, genome chips, and gene arrays) are a new technology that began to be used in the mid-1990s to analyze genome-wide patterns of gene expression within and among species (Gibson 2002, see also Chapter 7). Initial applications of microarrays involved gene discovery, disease diagnosis, drug discovery, and toxicological analyses comparing tissues or cells. However, this technology eventually may provide a new ecological tool by allowing entomologists to move beyond the one gene–one experiment situation.

Two basic types of microarrays are in use: in cDNA microarrays, small amounts of cDNAs are deposited at high density in spots on a glass slide or filter substrate. These cDNAs are hybridized to fluorescently labeled cDNA derived from two different RNA sources, and the ratio of the two signals at each spot reflects the relative levels of transcript abundance. The ratio is determined by the fluorescent colors obtained. The second type uses up to 20 microsquares of 25-mer oligonucleotides per gene, including perfect and mismatch pairs that will hybridize specifically or nonspecifically to a different part of the same transcript. Each square yields a different fluorescence intensity measure reflecting differences in GC content and folding of the RNA. In this case, the gene expression estimate is compared with measures from other chips (Mount 2001).

Microarrays allow the researcher to measure the relative quantities of specific mRNAs in two or more samples for thousands of genes simultaneously. When a sample containing bits of fluorescently labeled cDNA is added to the chip containing spotted samples of DNA or cDNA, the labeled samples (for example, one may be labeled green and one red) will anneal to the DNA on the chip that has complementary sequences. The chip contains known DNA sequences in a specific array. After complementary base pairing, the chip can be scanned and the colors will tell you which genes have hybridized with the sample DNA by complementary base pairing. Intermediate colors indicate those genes that were active in both samples. When one sample contains cDNA but the other sample does not, the sample will fluoresce either red or green. If the same genes are turned on in both insects, both red and green dyes will be present on the microarray "dot" and appear as yellow (Figure 13.1). Thus, DNA microarrays can provide information on the transcriptional changes (transcriptomics) that occur in specific cells, tissues, or whole organisms under specific conditions.

DNA microarrays provide a new way to analyze ecological interactions between plants and arthropod pests. For example, microarrays were used to assess the multiple transcriptional changes that occur after plants are attacked by insects (see the case study on insect–plant interactions described below). DNA microarrays might be useful for other ecological analyses. For example, scientists could use microarrays to evaluate the major events that occur during parasitism of insects by pathogens or nematodes or to evaluate the role of symbionts. Once the complete genomes of several strains of *Wolbachia* have been sequenced, microarrays could be used to learn more about the effects of *Wolbachia* on their hosts. DNA microarrays also might be applied to DNA profiling to characterize genetic differences among biotypes or closely related species (Gibson 2002). Limitations to the application of microarrays to ecological problems currently include their perceived expense and the unavailability of microarrays (Gibson 2002). As with DNA sequencing, however, the cost of microarrays is declining rapidly as more core facilities are devoted to this technology and as more genomes are sequenced.

Analysis of microarray data must deal with the challenge of comprehending and interpreting the resulting massive amounts of data. As with any new technology, quality control and adequate statistical methods must be developed and employed (Kerr and Churchill 2001a,b, Quackenbush 2001).

13.5.7. Microsatellites

DNA fingerprinting may involve the use of "microsatellite" or "minisatellite" sequences, which consist of arrays of up to several hundred simple sequence repeats (SSRs). In arthropods, these repeats most often consist of repeats of dinucleotides (AC, AT, AG), trinucleotides (AGC, AAC, AAT) or tetranucleotides (ACAT, AAAT, AAAC) (Toth et al. 2000). Microsatellites typically are scattered throughout the chromosomes of most organisms (Bruford et al. 1992, Estoup et al. 1993, Ashley and Dow 1994, Toth et al. 2000). Microsatellites are found both in protein-coding and noncoding regions. Microsatellites mutate at a high rate and are thought to play a significant role in genome evolution by creating and maintaining quantitative genetic variation (Kashi et al. 1997). For example, when microsatellites are found in promoter regions, they could influence transcriptional activity. The length of microsatellites could also affect protein–protein interactions during transcription.

Two models have been proposed to account for the high mutation rate in microsatellites. The first model, DNA polymerase slippage, assumes that replicating DNA strands transiently disassociate and then reassociate in a misaligned form, which will result in length

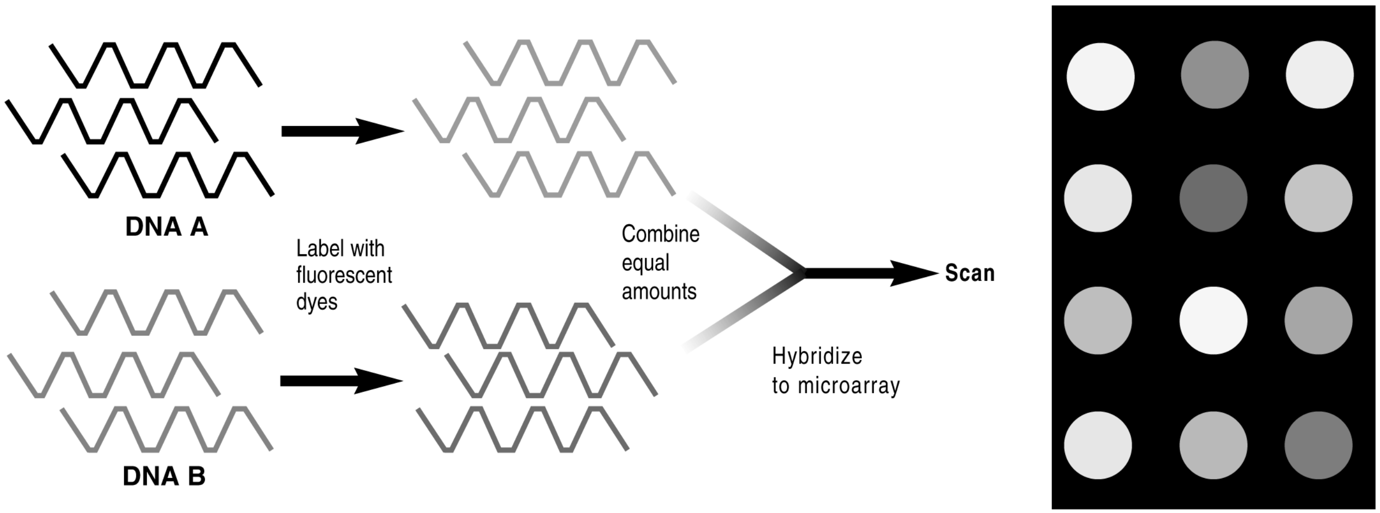


Figure 13.1. In some microarray experiments, cDNAs copied from the messenger RNAs of two different insects (A and B) are each labeled with a different colored fluorescent dye. The labeled DNAs are then combined in equal proportions and used as probes to hybridize with known DNA or RNA sequences on the microarray. After hybridization, the microarray is scanned, and the colors and their relative intensities are recorded. If, for example, DNA from insect A is labeled with red and DNA from insect B is labeled with green, when equal amounts of A and B DNAs are applied to the array the resulting color patterns indicate which genes are active in insect A and B. A red dot will indicate high levels of expression of the corresponding gene on the microarray in insect A (and no expression in insect B); a green dot in that area indicates high levels of expression of that gene in insect B (and no expression in insect A). If the same gene is equally active in both insects A and B, then the color will a blend of red + green, or yellow. If the mRNAs are present in a ratio other than 1:1, then other shades will be seen. [The dots on the right side of the illustration are shown in shades of gray here, but normally would be in color.]

and sequence variations. The second model involves unequal recombination to produce mutations in microsatellites. Understanding the evolution of microsatellites is considered important in using them for ecological analyses.

Microsatellite analysis results in a pattern of DNA bands on gels that resemble "bar codes" used to identify items in stores. DNA fingerprinting can be used to evaluate DNA variability at the individual and population level and was first conducted on humans and other vertebrates (Jeffreys 1987). The banding patterns produced often are specific to a particular individual (except for monozygotic twins), are inherited in a Mendelian manner, and are generally stable within an individual. Polymorphisms are visualized by hybridizing a labeled probe to genomic DNA that has been cut with a restriction enzyme and separated into bands on a gel by electrophoresis.

Microsatellite markers can identify individual insects or their progeny, evaluate kinship, resolve whether a mating has been successful, and reveal differentiation among closely related populations in the field (Burke 1989, Wang et al. 1999). Microsatellites could be useful in monitoring establishment and dispersal of specific biotypes, including those with low levels of protein variation such as parthenogenetic aphids or hymenopteran parasitoids (Table 13.1). Analysis of microsatellites has become a popular method for identifying high levels of genetic variability.

Unfortunately, microsatellite sequences differ in different organisms, even in closely related species. This means that microsatellite sequence data usually must be obtained for each species under study, making microsatellites relatively time consuming and expensive to develop. Furthermore, different taxonomic groups may exhibit differences in the ease with which microsatellites can be isolated (Neve and Meglecz 2000). For example, there were only five microsatellite studies published on Lepidoptera between 1997 and 1999, but 47 were published on Hymenoptera (Neve and Meglecz 2000); it is not clear whether equal efforts were expended or whether the disparity represents true differences in ease of isolation.

A variety of methods have been developed for isolating microsatellites (Table 13.2). Zane et al. (2002) review the methods and suggest a "fast and easy protocol which is a combination of different published methods." Their goal is to "provide a well established universal protocol," but they recognize that "completely new approaches [may] become available due to a better knowledge of microsatellite evolution combined with new technical advances." At present, a careful evaluation of the experimental strategy has to be carried out for each experiment (Zane et al. 2002).

Statistical issues associated with analysis of microsatellite markers were reviewed by Goodman (1997), Cornuet et al. (1999), Luikart and England (1999) and Balloux and Lugon-Moulin (2002).

13.5.8. RFLP Analysis

RFLP analysis can be used to analyze variation in both mtDNA and nuclear DNA (Table 13.1). Depending on which restriction enzymes are used and target sequences analyzed, extensive variation can be discerned. However, RFLP analyses require relatively large amounts of very clean DNA (which may not be obtainable from single individuals of small insects). The DNA must be digested, electrophoresed, blotted, and probed to detect the variation. Probes must be developed, either as consensus sequences from other species, or after cloning and sequencing species-specific DNA. Working with large numbers of individual insects is relatively time-consuming and expensive (Table 13.1).

Table 13.2. Selected Sequences and Databases Relevant to Insect Molecular Ecology

Primer or database type		Reference(s)[a]
Insect genomes		
D. melanogaster	Fly Base	FlyBase Consortium 2002, Gilbert 2002
Anopheles gambiae	Mosquito Genomics Web site	http://klab.agsci.colostate.edu
Apis mellifera	Expressed sequence tags	Whitfield et al. 2002, http://keckl.biotec.uiuc.edu/bee/honeybee_project.htm
Microarrays	Overview of methods and data analysis	Quackenbush 2001, Gibson 2002, www.gene-chips.com
Microsatellites	Overview of methods and analysis	Estoup and Angers 1998, Toth et al. 2000, Zane et al. 2002
Mitochondrial DNA	MitBASE	Attimonelli et al. 2000
	AMmtDB	Lanave et al. 2002
	Variable number of tandem repeats in mtDNA	Lunt et al. 1998
	Lepidopteran control region	Taylor et al. 1993
	PCR of long sections in 14 orders	Roehrdanz and DeGrugillier 1998
	12S rRNA	Hickson et al. 1996
	Complete sequences of 16+ insect and tick species	ncbi.nim.nih.gov:entrez
	PCR primers	Simon et al. 1994
	Alignments of mitochondria from 13 insect orders	Buckley et al. 2000
RAPD-PCR	Effective 10-mer primers in insects reviewed	Hadrys et al. 1992, Haymer 1994
Ribosomal		
5S rRNA	5S rRNA	Cullings and Vogler 1998, Szymanski et al. 2002
rRNA	Small-subunit rRNA	Wuyts et al. 2002

[a]A review of databases is published each January by the journal *Nucleic Acids Research*; search the most recent issue for updated locations and new databases. Also search the National Center for Biotechnology Information (NCBI) Web site for a variety of databases and data analysis methods (www.ncbi.nlm.nih.gov).

13.5.9. PCR-RFLP

A modification of RFLP analysis, called **PCR-RFLP**, eliminates many of the disadvantages of traditional RFLP analysis (Karl and Avise 1993, Table 13.1). If no probe is available, a genomic DNA library can constructed and clones isolated and sequenced. Alternatively, degenerate primers can be designed and the PCR products cloned and sequenced. Once sequences are available, allele-specific PCR primers can be designed. Subsequently, nuclear DNA is amplified by the PCR using these primers and the PCR product is digested with appropriate restriction enzymes. The cut DNA is visualized after electrophoresis by staining with ethidium bromide.

The advantage of PCR-RFLP is that DNA extracted from a single individual is sufficient (after PCR amplification) to provide bands that can be visualized. PCR-RFLP makes RFLP analysis suitable for studying individual specimens of very small species, requires no labeled probes, and is faster and less expensive than standard RFLP analysis. If consensus primers are available, then cloning is not required. For example, Simon et al. (1993) analyzed

mtDNA in 13- and 17-year periodical cicadas using standard primers for the COII-A6-A8-COIII segment (Simon et al. 1991). A disadvantage to PCR-RFLP is that the method requires two procedures and thus is more time consuming and expensive than allele-specific PCR.

13.5.10. RAPD-PCR

Hadrys et al. (1992) noted that RAPD-PCR is one of the most versatile methods available for molecular ecology because it can be used to determine biotypes or species identity, assess kinship, and analyze paternity. It can estimate genetic variation within populations and can be used to monitor colonization. RAPD-PCR is suitable for studying insects for which very little genetic information is available, requires only very small amounts of DNA, and can be used with very small insects. It is rapid and relatively inexpensive when compared to RFLP analysis, DNA sequencing, PCR-RFLP, DSCP, microsatellite, or microarray analysis (Table 13.1; see also Chapter 8).

Because RAPD-PCR uses short primers of arbitrary sequence (10-mers), it does not require the investigator to know the sequences of specific genes in order to develop primers for PCR. Haymer (1994) evaluated the sequences of various RAPD primers used on insects and listed 55 that had been particularly informative. RAPD-PCR primers sample both single-copy and repetitive DNA. Although the repeatability and reliability of RAPD-PCR can be problematic if care is not exercised, RAPD-PCR can provide inexpensive, repeatable, and useful data for some purposes (Penner et al. 1993, Edwards and Hoy 1993, MacPherson et al. 1993).

RAPD-PCR was used to analyze the amount of genetic variation in the parasitoids *Trioxys pallidus* and *Diglyphus begini* (Figure 13.2). DNA from individual *T. pallidus* was amplified using 120 different primers (Edwards and Hoy 1993). Of the 120 primers tested, 92 produced a total of 342 scorable bands, of which 118 exhibited presence/absence polymorphisms. *D. begini* was evaluated with 25 primers, and 17 produced a total of 51 scorable bands. The level of genetic variation detected was greater than any found in Hymenoptera using allozymes (Menken 1991, Packer and Owen 1992) and comparable to the amount of variation detectable with RFLPs. The bands considered "reliable" were inherited as dominant

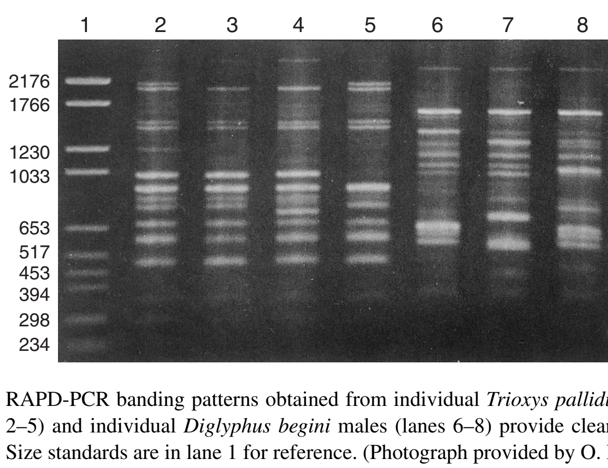


Figure 13.2. RAPD-PCR banding patterns obtained from individual *Trioxys pallidus* male wasps (lanes 2–5) and individual *Diglyphus begini* males (lanes 6–8) provide clear species differences. Size standards are in lane 1 for reference. (Photograph provided by O. R. Edwards.)

Mendelian traits (Figure 13.2). Because only small amounts of DNA are used for each RAPD reaction, multiple reactions can be conducted using the DNA from a single individual (Edwards and Hoy 1993, 1995). RAPD-PCR can be used to analyze population structure and gene flow and to monitor establishment and dispersal of particular biotypes (Hadrys et al. 1993, Black et al. 1992, Kambhampati et al. 1992, Edwards and Hoy 1995).

RAPD-PCR has drawbacks. RAPD bands are inherited as dominant traits in diplo-diploid species, and thus heterozygotes normally cannot be identified. In the haplo-diploid Hymenoptera, this difficulty is overcome by performing the analysis on haploid males only, or by testing the genotype of females by testing her male progeny (Edwards and Hoy 1993). RAPD-PCR conditions should be optimized for each species (Hadrys et al. 1992). Reliable bands must be identified by repeating reactions to determine which bands are consistent. RAPD-PCR is sensitive to the concentration of template DNA in the reaction, so reaction conditions must be optimized carefully and DNA extraction techniques must be consistent (Edwards and Hoy 1993). Primer quality is critical; fresh, undegraded primers should be used.

13.5.11. Sequencing

Sequencing provides large amounts of information about mtDNA and nuclear DNA. However, sequencing, because of time and economic constraints, can sample only a tiny fraction of the total genome. Also, despite reduction in the cost of sequencing, sequencing remains more expensive and time consuming than PCR-based methods. Sequencing of DNA amplified by allele-specific PCR requires information about sequences in order to develop appropriate primers. Sequencing has been used infrequently for large-scale population studies because of cost (Table 13.1) but is a useful tool for the other techniques such as developing allele-specific primers for the PCR. See Chapter 12 for methods of sequence analysis and the J. Felsenstein Web site for a variety of data analysis computer programs (www@evolution.genetics.washington.edu); Mount (2001) and Gibson (2002) provide detailed information for novices.

13.6. Analysis of Molecular Data

Molecular ecology is a developing and rapidly changing field of study (Ferraris and Palumbi 1996, Symondson and Liddell 1996), but is not yet quite mature. Methods of analyzing molecular population data are still being developed and improved. For reviews and overviews of methods, see Weir and Cockerham (1984), Slatkin and Barton (1989), Doolittle (1990), Lynch and Crease (1990), Weir (1990), Hoelzel and Dover (1991), Hoelzel and Bancroft (1992), Ferraris and Palumbi (1996), Bossart and Pashley Prowell (1998), Estoup and Angers (1998), Howard and Berlocher (1998), Rieseberg (1998), Schnabel et al. (1998), Templeton (1998), Davies et al. (1999a,b), Bohonak (1999), Cornuet et al. (1999), Goodnight and Queller (1999), Schwartz et al. (1998), Black et al. (2001), and Hewitt (2001). Computer software packages for molecular population genetic analyses are available from several sources (for example, Rozas and Rozas 1997, and the Felsenstein Web site: www@evolution.genetics.washington.edu/pub).

A variety of analyses can be conducted on molecular data to estimate parameters such as genetic diversity (heterozygosity and proportion of polymorphic loci), interpopulation diversity, genetic distance, effective population size, kinship, paternity, and the effect of

migration on population diversity, as described below. Fundamentally, molecular techniques produce one of two types of data: sequence data or allele frequency data (Table 13.1). Details of the statistical methods used are beyond the scope of this chapter. However, the references just cited can provide an entry to the extensive and growing literature.

13.6.1. Allozymes

The visualization and interpretation of allozyme data was reviewed by Pasteur et al. (1988) and May (1992). Allozyme data can be used to obtain gene frequencies (Hoelzel and Bancroft 1992).

$$p = (2\mathrm{N}_{\mathrm{AA}} + \mathrm{N}_{\mathrm{Aa}})/2\mathrm{N} \quad \text{and} \quad q = (2\mathrm{N}_{\mathrm{aa}} + \mathrm{N}_{\mathrm{Aa}})/2\mathrm{N}$$

where p is the frequency of the A allele, q is the frequency of the a allele, N is the total number of individuals in the sample, and N_{AA}, N_{aa}, and N_{Aa} are the number of individuals with AA, aa, and Aa genotypes, respectively. According to the Hardy–Weinberg rule, the proportion of AA individuals should be p^2, the proportion of aa individuals should be q^2, and the proportion of heterozygotes should be $2pq$ in an ideal population (infinitely large random-mating population) in which there is no selection, migration, or mutation. Such a population is in **Hardy–Weinberg equilibrium**.

Polymorphism (*P*) and **heterozygosity (*H*)** can be calculated for allozyme data (Hoelzel and Bancroft 1992). P is the proportion of polymorphic loci, and H is the proportion of heterozygous loci. When the population is in Hardy–Weinberg equilibrium, heterozygosity can be calculated from allele frequencies at a given locus by:

$$h = 1 - \sum x_i^2$$

where x_i^2 is the frequency of the ith allele at a given locus. The proportion of heterozygous individuals (H) is the average heterozygosity for all loci studied, so it is calculated as the mean of h over all loci.

Genetic distance between populations can be calculated using allozyme data. Most analyses of genetic distance assume that molecular genetic changes are accumulating gradually at a constant rate and that most changes are selectively neutral. This suggests that the genetic changes can be used to estimate the time of genetic differentiation within and between populations. When DNA variation is measured directly, the statistical analyses assume: 1) nucleotides are randomly distributed in the genome, 2) variation arises by base substitution, 3) substitution rates are the same for all nucleotides, and 4) all relevant bands or fragments can be detected and bands that comigrate but are different are not scored as identical (Hoelzel and Bancroft 1992). Although the first three assumptions usually are not valid, it is thought that small deviations from them will not alter the conclusions significantly.

The most commonly used method for analyzing genetic distance in populations by protein polymorphisms is that of Nei (1972). In two populations, X and Y, the probability that two randomly chosen genes at a single locus (j_k) are identical is:

$$j_x = \sum x_i^2 \quad \text{and} \quad j_y = \sum y_i^2$$

where x_i and y_i are the frequencies of the ith alleles at a given locus in populations X and Y, respectively. If there are two alleles at this locus with frequencies p and q, then

$$j = p^2 + q^2$$

The probability that a gene is identical at the same locus in populations X and Y is

$$j_{xy} = \sum x_i y_i$$

The normalized identity (I) between populations X and Y for all loci is:

$$I = J_{XY}/(J_x J_y)^{1/2}$$

where J_{XY}, J_x, and J_y are the arithmetic means of j_{xy}, j_x, and j_y, respectively, over all loci. **Nei's standard genetic distance (*D*)** between populations X and Y is then:

$$D = -\ln(I)$$

where the natural logarithm (ln) of *I* is taken to give a value that is 0.0 for genotypes that are completely dissimilar. The relationship between *D* and time (*t*) is:

$$t = 0.5aD$$

where *a* is the average rate of detectable change per locus per year.

Interpopulation diversity using allozyme data are usually measured using the **coefficient of gene differentiation (G_{ST})**. G_{ST} is derived by estimating the average similarity within and between populations. G_{ST} is an extension of Wright's correlation (F_{ST}) between two gametes drawn at random from each subpopulation. The coefficient of differentiation is:

$$G_{ST} = (H_T - H_S)/H_T$$

where H_S is the average gene diversity within populations, and H_T is the interpopulation gene diversity.

13.6.2. Microsatellites

Microsatellites can identify multiple loci or single loci in individuals (Zane et al. 2002). Multiple-locus DNA fingerprinting uses satellite sequences scattered throughout the chromosomes to produce a series of bands that are often specific to an individual insect. Microsatellites may detect so much variation within populations that it is difficult to analyze them unless inbreeding has occurred in the population under study so that some of the variability has been lost. DNA fingerprinting using microsatellites can be done with the PCR using specific or consensus primers (Kirby 1990).

Population estimates of allele and genotype frequencies can be tested for correspondence to Hardy–Weinberg equilibrium conditions (Bruford et al. 1992), and genetic differentiation can be calculated from microsatellite data (Goodman 1997, Goodman and Queller 1999). The high level of variation detected by satellite data makes it feasible to test for paternity and to conduct studies of variability within both sexual and clonal populations (Brookfield 1992). Variation and genetic distance also can be calculated (Hoelzel and Bancroft 1992). Single-locus DNA fingerprinting is easier to analyze because there are fewer bands, but advances are being made in analysis of multiple microsatellite loci (Estoup and Angers 1998).

13.6.3. RAPD-PCR

RAPD-PCR bands are considered as dominant loci in diplo-diploid organisms, and scored as present or absent (Hadrys et al. 1992). Kambhampati et al. (1992) discussed appropriate statistical methods for analysis of data. It appears that RAPD-PCR loci can be used to determine paternity, kinship, and hybridization, as well as to estimate population heterozygosity and effective population size, identify biotypes and cryptic species, and measure genetic distance between populations and interpopulation diversity (Table 13.2).

13.6.4. RFLPs

Visualization and interpretation of RFLP data were described by Aquadro et al. (1992) and Dowling et al. (1990). Restriction patterns can be compared either by the lengths of the fragments or by comparing actual restriction sites. Restriction patterns can be classified as haplotypes, and a measure of diversity can be derived as a function of the frequency of the different haplotypes (Hoelzel and Bancroft 1992). The term **haplotype** is a contraction of haploid and genotype and describes the combination of linked alleles in a cluster of related genes. Likewise, genetic distance is measured as an estimate of the number of base substitutions per nucleotide separating the two populations. Interpopulation diversity (G_{ST}) can be estimated in a manner similar to that for allozyme data, but gene identities must be estimated from RFLP patterns. RFLP data also can be analyzed as changes in base pairs if the assumption is made that each change in restriction pattern is caused by a change in a single base pair.

13.6.5. Sequencing

DNA sequence data are analyzed by computer programs to determine the best alignment (Doolittle 1990, Gribskov and Devereux 1991, Gibson and Muse 2002). The identity of two sequences is compared on the percentage of shared bases. Deletions and insertions are usually scored as a single change regardless of length (Hoelzel and Bancroft 1992). As with proteins or RFLP data, nucleotide diversity, genetic distance, and interpopulation diversity can be estimated (Hoelzel and Bancroft 1992). See also Chapter 12 for details on DNA sequence analysis methods.

13.7. Case Studies in Molecular Ecology and Population Biology

An extensive and growing literature published in a variety of journals and books makes it impossible to provide a comprehensive overview of the effect that molecular methods are having on theoretical and applied insect ecology. Thus, several case studies will be presented to illustrate applications of several different molecular tools and their statistical methods.

13.7.1. Genetic Variability in the Fall Armyworm

The fall armyworm, *Spodoptera frugiperda*, is a polyphagous lepidopteran "species" that attacks more than 60 varieties of plants, particularly corn and bermuda grass. This migratory pest overwinters in southern Florida, southern Texas, and the Caribbean, but disperses

north during the late spring and early summer. The fall armyworm has exhibited anomalous differences in tolerance to pesticides and other life-history traits that are important in pest management practices. It appears that it is, in fact, undergoing "incipient speciation" (Pashley 1986).

The undetected presence of cryptic species can have practical significance for pest management programs. Pashley (1986) showed that the two "species" are each associated with different host plants: one with rice and bermuda grass and the other with corn. The two occur sympatrically, exhibit a high level of reproductive isolation, and are physiologically adapted to the different host plants; the physiological differences are genetically based (Pashley 1988).

Allozyme and RFLP analysis of mtDNA indicated the two types could be distinguished and that there is a near absence of gene flow (Pashley 1989). Lu et al. (1992) reported RFLP differences in genomic DNA in the two types after analyzing six colonies with 22 different markers. Lu et al. (1994) also found that repeated DNA sequences (microsatellites) in the two populations differed. The extent to which the sympatric populations of rice and corn strains of *S. frugiperda* interbreed is problematic (Pashley 1986). Some data suggest there is a unidirectional behavioral barrier to interstrain mating, but other data do not support this (see review by McMichael and Pashley Prowell 1999).

To answer whether these populations interbreed, several different molecular markers were used because allozymes and mtDNA and nuclear DNA markers did not provide sufficient resolving power to discriminate between the alternative hypotheses: low frequencies of hybridization versus expected genetic overlap between two closely related populations. AFLP-PCR was evaluated to determine if this tool might uncover unique genetic markers in each of the two populations, which would allow hybridization to be detected more readily.

McMichael and Pashley Prowell (1999) used ten AFLP markers to compare the two populations of *S. frugiperda*. The AFLP data identified two populations that matched up with the majority of individuals from one or the other of the host-associated strains, as defined by habitat and mtDNA. Unfortunately, not all individuals could be assigned to the "rice" or "corn" populations. To date,

> no pair of markers shows complete congruence with each other or host of origin. In other words, allozyme or mtDNA genotypes characterizing one strain can occur in individuals collected on the other strain's host. Individuals on a single host can contain an allozyme genotype characteristic of one strain but a mtDNA genotype of the other (McMichael and Pashley Prowell 1999).

These results cannot discriminate between the alternative hypotheses: sharing of alleles because the variability in the common ancestor of the two strains did not become fixed during their divergence, or interstrain hybridization. Thus, the goal to identify diagnostic, and unique, AFLP markers failed, and McMichael and Pashley Prowell (1999) concluded that future studies, in which AFLP data are combined with mitochondrial markers and allozymes, might resolve the hybridization question.

The molecular data, in combination with other data, clearly have shown that the two populations are different and explain a longstanding concern of practical pest management importance. Whether these populations are called species, incipient species, or host races is a judgment call that is based on whether the scientist is a "splitter" or a "lumper."

AFLP genotyping or fingerprinting is emerging as a useful tool for assessing genetic diversity, relatedness, population structure, and phylogenetic relationships (Mueller and Wolfenbarger 1999). Mueller and Wolfenbarger (1999) noted that AFLPs can be more reliable than RAPD markers, can be more user-friendly than RFLPs and microsatellites,

and, although not a panacea for molecular ecology, will probably replace several techniques in population genetics, fingerprinting, and systematics studies.

13.7.2. Population Isolation and Introgression in Periodical Cicadas

Molecular markers have been used to resolve the evolutionary origins of species of periodical cicadas (*Magicicada*). The biology, ecology, and evolution of periodical cicadas are complex and unusual (see Marshall 2001 for a review). Periodical cicadas feed underground on roots in the deciduous forests of the eastern United States for either 13 or 17 years and emerge in very large numbers as adults to mate and deposit eggs nearly every year in some part of the range. The immense populations, sometimes as large as 1.5 million individuals per acre, that emerge in the same year are called broods. This synchronized emergence may have evolved because the large numbers allow most of the individuals to escape predation at a particular location and the long life cycles may prevent predator populations from synchronizing with the local emergences.

In the Mississippi Valley and southern United States, the life cycle of three *Magicicada* species is 13 years, whereas it is 17 years for three species in the north and west. Each species appears most closely related to another with the alternative life cycle, so that there are "species pairs" (13 paired with 17). This pattern suggests that speciation in *Magicicada* may involve a combination of geographic isolation and life-cycle changes that create reproductive isolation by changes in emergence patterns.

Thirteen of the possible 17 broods of the 17-year cicada *M. septemdecim* and three of the possible 13 broods of the 13-year cicada *M. tredecim* have been identified, and their emergence patterns have been identified and monitored (Marshall 2001). Although most broods emerge as scheduled, small numbers of a brood may emerge "out of step" with their cohort, which has created problems in understanding the species status of some broods. Some portions of 17-year broods appear to have accelerated their emergence by 4 years in certain sites, and both 13- and 17-year broods may emerge in the same geographic region. The reason for the 13- and 17-year cycles may be because the life cycle of the 17-year cicada includes a 4-year inhibition (diapause) of early nymphal growth. It was suggested that if this inhibition were eliminated, the 17-year brood could emerge after only 13 years. Because it appears that a single gene controls this aspect of the life cycle, a relatively simple genetic change could have a large effect. An alternative hypothesis for the change in brood duration is that a 17- and 13-year brood emerged together in 1868 in Illinois and "hybridized," which resulted in a population that subsequently emerged every 13 years.

Were the cicadas newly emerging after 13 years derived by hybridization or by loss of a 4-year diapause? To resolve this intriguing evolutionary and ecological question, Martin and Simon (1988) analyzed the abdominal sternite color, the frequency of allozyme polymorphisms, and mtDNA of the "hybrid" Illinois population. The data indicated that mtDNA in the 13-year Illinois brood is like that in the adjacent 17-year brood and distinct from that of the neighboring 13-year brood. The new brood is like the 17-year brood in abdominal color and frequency of PGM (protein) polymorphism. These results are consistent with the hypothesis that the two populations hybridized and that the 13-year life cycle trait is dominant.

Nevertheless, Martin and Simon (1988) rejected the hybridization hypothesis. They pointed out that most cases of hybridization involve narrow zones in which the species come into contact, yet this new periodical cicada population occurs over a large area. They also noted that the complete elimination of one of the mtDNA genotypes throughout

the entire region would require *extremely* strong selection because there have been only eight generations since 1868 upon which selection could have occurred. Furthermore, if hybridization occurred, intermediate phenotypes should have occurred, but did not. As a result, the findings were interpreted as evidence for a widespread life-cycle switch, *without hybridization*, in which a large number of 17-year cicadas (*M. septemdecim*) underwent a 4-year acceleration in development to become 13-year cicadas in northern Arkansas, Missouri, Illinois, and southeastern Iowa (Martin and Simon 1988). After 1868, a portion of the progeny of brood X in these areas had a permanent 4-year acceleration in development rate. Some cicadas in the region did not switch life-cycle length after 1868 and continued to emerge after 17 years. Martin and Simon (1988) suggest the switch in life-cycle length could have been triggered by environmental causes, perhaps by high-density populations. This change in life-cycle length would disrupt gene flow and initiate genetic divergence in the formerly unified 17-year cicada population. Furthermore, the individuals that switched their life cycles from 17 to 13 years now emerge synchronously with a previously isolated brood that emerges every 13 years (a *M. tredecim* brood). Gene flow between these previously isolated broods could occur if they can mate and produce viable progeny.

Further analyses of the populations were carried out by Marshall and Cooley (2000). They found that the two 13-year sympatric cicada populations in the midwestern United States exhibit song differences and thus are unlikely to interbreed in the field. As a result, they described the "new" 13-year population as a previously undescribed species called *Magicicada neotredecim*. *M. neotredecim* and *M. tredecim*, both 13-year cicadas, overlap geographically and, since 1868, their broods overlap chronologically. Furthermore, *tredecim* and *neotredecim* populations differ in abdomen coloration and mtDNA. Such traits in *M. neotredecim* are not consistently different from the sympatric populations of the 17-year *M. septemdecim*, and thus *neotredecim* appears most closely related to this geographically adjacent population with a 17-year life cycle. Marshall and Cooley (2000) suggest this is evidence that speciation in *Magicicada* involves temporal isolation.

Simon et al. (2000), in a companion article, support the conclusion that *M. neotredecim* is a new cryptic species, and present mtDNA data to support the conclusion that a life-cycle switch occurred to produce two overlapping 13-year cicada lineages. Furthermore, the genetic evidence suggests that assortative mating is taking place in the area where the two populations of 13-year cicadas overlap. Thus, Simon et al. (2000) propose two possible scenarios for the evolution of the *septemdecim-tredecim-neotredecim* species (Figure 13.3). In both models, the ancestor is a cicada with a 13-year life cycle, which gave rise to a 13-year lineage (*tredecim* lineage) and to a new lineage that had a 4-year extension (17-year *septemdecim* lineage). In one scenario, *neotredecim* evolved from *septemdecim* via a single 4-year life-cycle reversion (instantaneous speciation) and eventually split into the two current broods (Figure 13.3A). In a second scenario, the *septemdecim* lineage gave rise to two independent 17- to 13-year life-cycle reversions, resulting in two *neotredecim* broods (Figure 13.3B). There are no genetic or behavioral data to distinguish between these two hypotheses at this time.

Simon et al. (2000) addressed the question as to whether the 13- and 17-year cicadas should be called different species. The concept of "species" varies among different systematists, but Simon et al. (2000) support the thesis that species should be designated as soon as they are distinguishable if it is likely that they will remain extant and isolated long enough for reproductive isolation to be developed. Marshall and Cooley (2000) presented behavioral data and field observations to suggest that important differences do exist

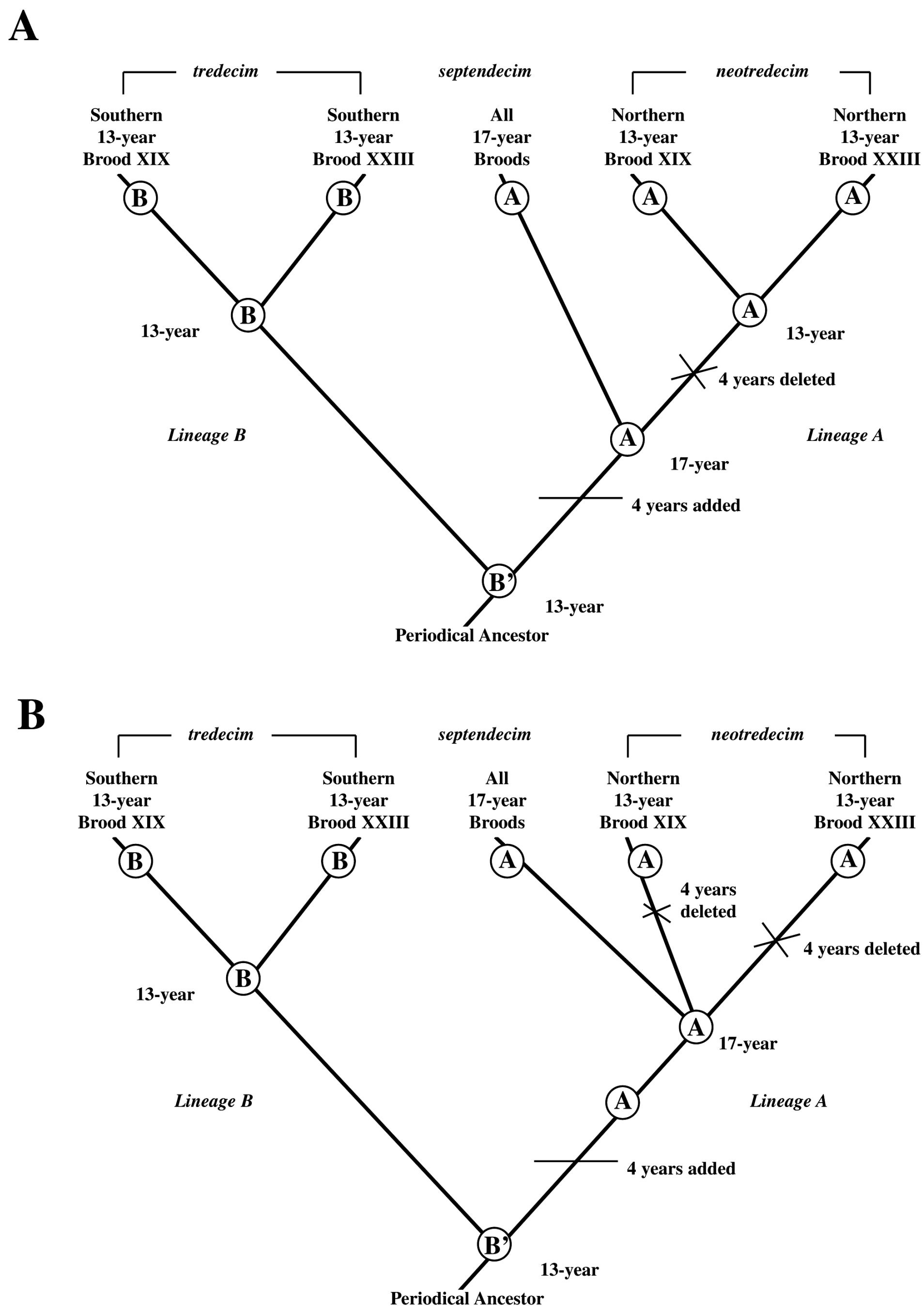


Figure 13.3. Two hypotheses for the formation of 13- and 17-year lineages of the *Magicicada septendecim, tredecim*, and *neotredecim* group. (A) The evolution of *neotredecim* involved one evolutionary event labeled (2). Alternatively, (B) the evolution of *neotredecim* involved two separate events (labeled 2 and 3). The models assume that *tredecim* populations were ancestral in both cases. (Modified from Simon et al. 2000.)

in male song and female response, supporting the separation of *neotredecim* and *tredecim* as species. They also argue that *neotredecim* is unlikely to revert to a life cycle that is synchronized with any extant 17-year brood. Despite behavioral and genetic similarities between the 17- and 13-year broods, it is likely that reproductive isolation will be achieved between *neotredecim* and *septemdecim* because of the differences in brood timing. Simon et al. (2000) summarize their hypothesis as follows: "an initial allochronic event separated lineage A and B periodical cicadas and a second allochronic event placed them in secondary contact."

Cooley et al. (2001) carry the story forward and predict that the newly described *M. neotredecim* inhabits midwestern habitat that will be unsuitable for its survival during the next "ice age." They argue that the most southern populations of *neotredecim* have the greatest "likelihood of colonizing a refugium during the next glacial cycle . . . then after the next glacial retreat all undisplaced *M. neotredecim* will have gone extinct, and the pattern of character displacement linking *M. neotredecim* to an allochronic speciation event in the *M. septemdecim* lineage will have been erased." Thus, the *Magicidada tredecim–septemdecim–neotredecim* story supports the hypothesis that instantaneous speciation can occur by shifts in reproductive timing (allochrony), although the precise mechanism by which the 4-year shift occurred remains unknown.

13.7.3. Eradicating Medflies in California?

The Mediterranean fruit fly, *Ceratitis capitata*, is an immensely destructive pest of agriculture. It is a native of sub-Saharan Africa, but invaded the Mediterranean basin, portions of Central and South America, Hawaii, and Australia during the past 100 to 200 years. *C. capitata* is able to feed on more than 200 species of host plants and can survive in a variety of climates (USDA-APHIS 2002). Females deposit up to 1000 eggs in fruits (including peaches, pears, plums, apples, apricots, avocados, citrus, cherries, figs, grapes, guavas, kumquats, loquats, nectarines) or vegetables (peppers, tomatoes) and the resulting maggots eat the fruits or vegetables, leaving them mushy and infested with mold. Medflies can develop from egg to adult in 21 days, so populations can increase exponentially in favorable climates.

C. capitata is a quarantined pest, meaning that when it is detected in the United States efforts are made to eradicate it (USDA-APHIS 2002). Eradication efforts are justified because: this pest significantly increases production costs; pesticide applications to suppress Medfly can disrupt biological control of other pests; and fruits grown in Medfly-infested regions cannot be exported to Medfly-free areas, thereby affecting national and international trade. Furthermore, establishment of Medfly would create serious pest problems in backyard gardens and orchards.

Eradication efforts involve surveys, regulation, and control. Surveys are conducted by the USDA-APHIS and the States by placing Medfly traps in high-risk areas, especially near international airports and seaports. If an infestation is found, additional traps are placed to determine the extent of the infestation; unfortunately, the traps are not 100% efficient and very low populations can be missed. Control methods include application of aerial and ground bait sprays, release of large numbers of sterile flies (SIRM method), and application of pesticides to the soil under infested trees to kill larvae as they enter the soil to pupate and the adults as they later leave the soil. Movements of host plants out of the infested area is prohibited and, in some cases, infested fruits may be picked and destroyed.

13.7.3.1. The Controversy

Since 1975, California has grappled with a controversial problem. The periodic and repeated appearance of Medflies in traps, especially in the Los Angeles basin, has raised questions as to whether California can export agricultural products as a "Medfly-free state" and whether the eradication efforts have been successful (Carey 1991, Abate 1993, Carey 1996a,b, Myers et al. 2000). This problem has received national and international attention due to the large amounts of money involved; California exports more than $1 billion in fruits and vegetables to other countries, and approximately $1 billion in produce is sold domestically and in neighboring countries. Estimates of damage to California's economy due to direct damage to agriculture and to related jobs are enormous (Abate 1993, CDFA 2002, USDA 2002).

> If California fruits were quarantined from all foreign markets because of Medfly infestation, the state would suffer a loss of 35,000 jobs and experience reductions in output of $3.6 billion . . . The worst case would be if all other states also embargo California fruits. This could result in more than 132,000 jobs lost, $13.4 billion lost in economic activity and more than $3.6 billion in lost income to California families (CDFA 2002).

Because southern California is a gateway to Latin America and Hawaii, where Medfly is endemic, there is a constant risk of Medfly introductions. Medflies could be transported by millions of international travelers, commercial fruit smugglers, and mailed packages. For example, more than a million passengers enter the Los Angeles airport annually from Hawaii, where Medfly is endemic; if only one visitor in a thousand illegally transported infested fruit from Hawaii, there could be 1000 opportunities annually to create an infestation in California (CDFA 2002).

The first Medfly infestation in California was detected in Los Angeles in 1975; sterile Medfly releases and ground applications of malathion and bait were carried out, and the infestation soon was declared eradicated. In 1980, a total of 180 Medflies were found and eradicated. Few Medflies were found between 1982 and 1987 in California. However, the questions really began in the late 1980s: in 1987, 43 were found in Los Angeles County, and eradication efforts resulted in aerial sprays followed by the release of 1 million sterile Medflies per square mile per week. In 1988, 54 Medflies were found in two locations in Los Angeles County, which was treated with an aerial spray and releases of 1 million sterile Medflies per square mile per week. In 1989 and 1990, a total of 304 Medflies were found, which resulted in an eradication program that included 21 treatments in four counties covering 536 square miles. One wild Medfly was found in 1991 in October, and trapping located an additional 24 flies; eradication efforts included trapping, ground application of malathion and bait, and sterile Medfly releases over a 26-square-mile area using 30 million sterile flies per week between October 1991 and August 1992. During 1992 and 1993, 202 wild Medflies were found in the Los Angeles basin, and eradication efforts employed trapping, ground applications of malathion and bait, and sterile releases. In 1993, 400 flies were found, leading to eradication efforts that included eight aerial applications of malathion and bait plus mass trapping. During 1994, a "basin wide" sterile Medfly release program was initiated over a 1464-square-mile area; 250,000 sterile flies per square mile per week were released over the entire area, and 250,000 sterile flies were released, as well, in the areas where 73 wild flies were found in 1994; this effort was concluded in March of 1996.

In 1996, a "preventative" release program was initiated over a 2155-square-mile area in four counties to "prevent the development of Medfly infestations and to limit the geographic size of any that manage to start" (CDFA 2002). The releases involved at least 125,000 sterile Medflies per square mile per week and an additional 125,000 sterile flies over a high-risk

area in central Los Angeles. No Medflies were found during 1996. During 1997, 24 wild Medflies were found in Los Angeles County, which led to increased efforts, including ground sprays, fruit stripping, soil drenches with pesticides under infested trees, and sterile fly releases that were increased to 500,000 per square mile per week within a 9-square-mile area around each infested site.

Between 1975 and 1990, aerial spraying of malathion in baits had been used to kill the Medfly but, after considerable public concerns were raised about potential negative environmental and health effects, aerial application of this pesticide was banned in California. In the early 1990s, agricultural experts and entomologists hotly debated whether the eradication methods were effective. One camp believed that a combination of releasing millions of sterile Medflies, plus implementation of effective quarantines and the use of attractants and traps to delineate the infestation zone and to monitor the effectiveness of the SIRM program, was effective. Others believed that eradicating the Medfly from the Los Angeles basin was nearly impossible and that these methods were failures.

One of the significant questions in the debate is whether the Medfly was a "permanent resident" of California, especially in the Los Angeles area, or whether the ongoing outbreaks are the result of additional independent invasions (Carey 1996a,b, Headrick and Goeden 1996, Myers et al. 2000, Figure 13.4). This issue has been hotly debated and created considerable animosity. Until 1990, scientists and government officials assumed that each Medfly outbreak originated from flies that arrived from another country, hitchhiking within agricultural products imported or smuggled into California; many believed that the source of the invasions was Hawaii. The other group believed, however, that the Medfly had become permanently established and was being maintained at undetectable levels, due in part to the inability of Medfly traps to detect very low-density populations.

Molecular methods have been used in attempts to answer three questions in the Medfly story: Are species-specific diagnostics useful for ecological studies and quarantine procedures? What is the colonization history and population structure of Medflies as they moved out of Africa and into California? Is the Medfly permanently established in the Los Angeles basin?

13.7.3.2. Species-Specific Diagnostics

An important component in dealing with invasion problems is to be able to rapidly and reliably identify the introduced species. Within the Diptera, the family Tephritidae contains many important agricultural pests. Of the more than 4000 species of tephritids in infested fruits, 250 are considered pests (Armstrong et al. 1997). Unfortunately, it is difficult to identify immature tephritids, so rearing is often required to obtain accurate identifications, which can cause a significant delay in a quarantine or eradication program.

Using RAPD-PCR, Sonvico et al. (1996) found that it is possible to discriminate between immature stages of Medflies and *Anastrepha fraterculus*. Haymer et al. (1994) identified unique repetitive DNA probes that they used in slot blots or squash blots to discriminate between eggs or larvae of three tephritid species. The squash-blot procedure used a non-radioactive hybridization and detection method, making it simple and rapid to carry out and potentially allowing rapid identification of infested fruits at the earliest stage during quarantine and eradication procedures.

AFLP-PCR was used by Kakouli-Duarte et al. (2001) to discriminate between *C. capitata* and *C. rosa*. A species-specific repetitive marker was cloned and used as a probe for genomic dot-blot hybridizations; the probes were sequenced and primers were developed.

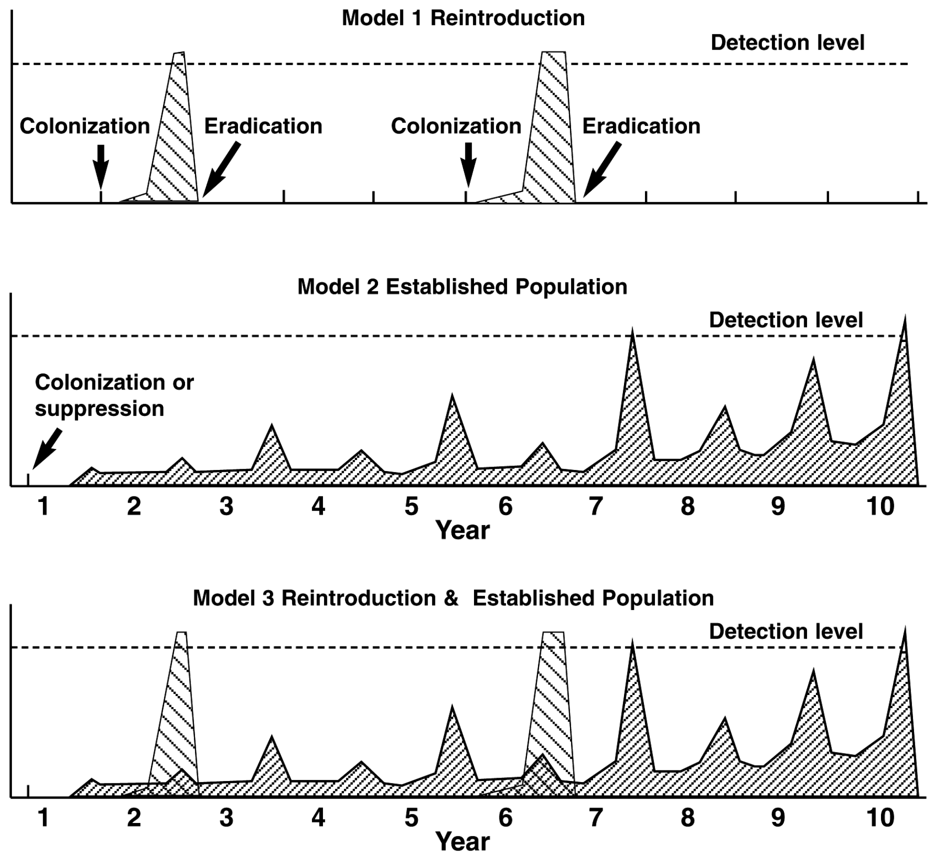


Figure 13.4. The top two models represent the alternative hypotheses for the presence of Medfly populations in California in the 1990s. The lowest "combination" model illustrates a situation in which both establishment and reintroductions are occurring. A more complex situation similar to this "combination model" appears consistent with the molecular data. The top two models were redrawn from Carey (1991).

AFLP-PCR products from *C. capitata* electrophoresed on polyacrylamide gel revealed diagnostic bands after silver staining (Kakouli-Duarte et al. 2001). However, the probe produced faint bands with DNA from *Bactrocera cucurbitae* and *B. oleae*, suggesting that the repetitive DNA fragment exists in low copy number in them. The number of outgroup tephritids tested was limited, so the probe may only be useful to discriminate among a few species. However, the authors did investigate the method's sensitivity and found that, because the probe sequences are found in multiple copies in the genome, sufficient DNA could be extracted only from wings or legs of *C. capitata* yet yield positive results.

Ribosomal ITS1 polymorphisms were investigated in *C. capitata* and *C. rosa* to provide species-specific probes and to investigate the differences in size among different populations of each species (Douglas and Haymer 2001). Recall that insects have multiple ribosomal genes (rDNA) and that the noncoding ITS (internal transcribed sequences) tend to vary sufficiently to allow discrimination of lower level taxonomic groups. ITS1 sequences were isolated from a genomic *C. capitata* library and sequenced; sequences from *C. capitata* and from *Drosophila melanogaster* were aligned to identify conserved sequences, and primers

were designed to amplify across the variable ITS1 regions. No obvious size variability in the PCR products was found among *C. capitata* from Hawaii, Guatemala, Spain, Greece, Costa Rica, and Peru, and sequencing of some of these ITS1 regions revealed only single nucleotide changes, with length variation ranging only from 829 to 832 bp. Thus, ITS1 sequence data would not allow a researcher to discriminate among these populations, but they do allow *C. capitata* to be identified. The same primers amplified ITS1 sequences from two populations of *C. rosa* and produced products that were 717 bp and 930 bp long, which might be used to distinguish *C. rosa* individuals from Kenya from those from South Africa (Douglas and Haymer 2001) if additional populations from each geographic region can be tested and the utility of the length variation is confirmed.

Armstrong et al. (1997) reported efforts to identify tephritids for quarantine purposes in New Zealand using 18S and 18S + ITS regions of ribosomal DNA. The 18S and 18S + ITS regions were amplified from larval DNA by the PCR, and 19 species in four genera were evaluated. Restriction analysis of the 18S product provided poor resolution, even at the generic level. Digestion of the 18S + ITS PCR product generated 13 diagnostic haplotypes using four restriction endonucleases. Six of ten *Bactrocera* species could not be diagnosed separately with this method despite analyzing the effects of 22 restriction enzymes. However, all six species are high risk with respect to their likely establishment in New Zealand, so a diagnosis of suspicious larvae as *Bactrocera* would result in the same response by regulatory authorities.

The studies just described provide an overview of some of the approaches that could be taken to use molecular methods to identify *C. capitata* immature stages and to discriminate *C. capitata* from other tephritids. However, additional research would be required before these approaches could be used in specific quarantine procedures, and none are useful for all objectives. For example, if a DNA probe is intended to identify *C. capitata* immatures when tephritid larvae are found in infested fruit in border inspections, then the probe should be validated against all potential species that might be introduced (recall that approximately 250 tephritids are considered pests). However, it may not be relevant to identify larvae to the species level; rather it may be relevant only to confirm that the larvae are from one of the tephritid pest genera because all would be quarantined pests. In addition, a useful probe ideally would be accurate; how often do false positives and false negatives occur under real-world conditions with the test method employed? In some ecological studies, it might be relevant to discriminate between immatures of two, or a few, tephritid species, which would be much easier, because the test needs to be validated only against these species.

13.7.3.3. Geographic Origin of Medfly Populations

One of the most difficult questions to answer involves efforts to resolve the geographic origin of Medfly populations in new environments. Several studies have been conducted using RAPD-PCR, enzyme analysis, nuclear introns, mitochondrial DNA, RFLP-PCR, and microsatellites. Some projects will be described briefly to illustrate that each method has strengths and weaknesses in answering a specific question.

Regulatory agencies often want to know where a pest population came from because it might allow them to prevent future invasions if they know where to invest their inspection and detection efforts. It also is relevant if sterile Medflies are to be released in an eradication program; potentially, sterile Medflies of one "type" might not mate with an invasive wild population if there are sufficient genetic differences in behavior or other premating isolation mechanisms. With regard to the California Medfly eradication efforts, the origin or

genetic makeup of the different Medfly invasions could answer the question as to whether the expensive eradication efforts are successful.

Malacrida et al. (1996) used enzyme electrophoresis to analyze genetic similarities among 11 tephritid species. A later study focused on tracking the colonization of *C. capitata* throughout the world using samples from 17 populations (Malacrida et al. 1998). Variability at 26 polymorphic enzyme loci revealed "that the geographical dispersal of medfly from its ancestral source area (East Africa) is associated with a great reduction in variability. The pattern of decreasing variability occurs at two regional levels: in the African-Mediterranean region where the differentiation is gradual, and in the Latin American-Pacific region where some ancestral variability is still present as a consequence of recent colonization" (Malacrida et al. 1998). The molecular data confirm that the name "Medfly" is inappropriate—because the ancestral home of *C. capitata* is Africa, it ought to be called "Africafly." Malacrida et al. (1998) concluded that the "population genetic changes observed in the species range are consistent with both the chronology and the historical circuitous course of the medfly colonization process." Thus, the molecular data are congruent with what is known about its movements out of Africa.

13.7.3.4. Is the Medfly Established in California?

A critical question regarding the success of eradication in California is: "Is the Medfly permanently established in California?" The question has been approached by assuming that independent introductions of Medflies from different geographic sources would result in populations with unique genetic markers; if each invasive population had different markers, it would be evidence that each invasion is independent. By contrast, if the markers found in the California populations during the different "invasions" were the same, the conclusion could be that it is more likely that a single Medfly population is established in California. However, an alternative explanation for Medflies having the same markers is that multiple invasions occurred from a particular geographic source. Unfortunately, it could be impossible to exclude this possibility.

Obtaining definitive data to discriminate between populations depends on having markers that are diagnostic. These markers need to be validated with large samples of flies from different geographic regions to confirm that the differences detected in the preliminary screening hold up when larger samples of, potentially more diverse, flies are sampled. The task of identifying appropriate genetic markers has engaged a number of researchers and considerable funds over the past few years.

13.7.3.4.1. RAPD-PCR and Allozymes

Early attempts to discriminate between different geographic populations of the Medfly used RAPD-PCR markers (Haymer and McInnis 1994) or enzyme electrophoresis (Malacrida et al. 1996) and compared RAPD-PCR and enzyme electrophoresis data (Baruffi et al. 1995). As expected, RAPD-PCR revealed larger amounts of genetic variation than enzyme electrophoresis data (Baruffi et al. 1995). The complete mitochondrial genome of the Medfly was sequenced, and different populations were found to exhibit genetic differences that are potentially useful for developing diagnostic tools (Spanos et al. 2000).

13.7.3.4.2. PCR-RFLP

A PCR-RFLP method was used by He and Haymer (1999) to compare variation in intron sequences of the glucose-6-phosphate dehydrogenase gene among different Medfly populations. Five alleles of this locus were found in 26 populations of *C. capitata*, and two

restriction enzymes were used in successive digestions of the PCR products to document genotypes and allele frequencies. This approach involved amplifying intron sequences from individuals from various populations using primers designed from cDNA sequences in GenBank. Allelic variants were identified by cloning and sequencing the PCR products, and restriction-site changes were identified. The restriction-site data allowed He and Haymer (1999) to develop a diagnostic test that did not require sequencing of PCR products. The data were analyzed using a principal coordinate analysis and analysis of molecular variation (AMOVA) to quantify the distribution of genetic diversity in a hierarchical manner. For some of the invasive sites, populations that "are probably acting as sources of origin" were identified (He and Haymer 1999). The five alleles tended to be associated with populations from different geographic regions: A1 was most common and was found in all populations surveyed, so it is not informative. A2 and A3 were widespread in samples from Greece, but only one allele tended to be prevalent in other samples (A2 was prevalent in samples from Guatemala, Peru, Florida, and southern California; A3 was prevalent in samples from Argentina). Hawaiian populations showed substantial frequencies of A4, but A4 was rare in other populations.

He and Haymer (1999) concluded that "the invasive population from northern California appears similar to populations from Argentina and Costa Rica." From southern California, three of the infestations (1992–1994) are clustered with populations from Guatemala, suggesting "that Guatemala is a possible source of origin of these flies." The 1997 southern California infestation is "well separated from either of the two previous groupings . . . these results suggest it is not appropriate to group or depict all of them as homogeneous" (He and Haymer 1999). They concluded that "the extreme separation of the Hawaiian populations from these California (and Florida) infestations also suggests that Hawaii can be considered a very unlikely source." Furthermore, "Samples from the California 1997 infestation are also well separated from all other populations, suggesting that none of the worldwide populations sampled here can be considered likely sources." Thus, "the multiple infestations detected within California in recent years are not likely to represent a single, homogenous population that is similar to the 'established' populations seen in Guatemala or Argentina" (He and Haymer 1999). The authors noted that additional surveys of these markers in populations from other regions of the world are desirable in order to improve resolution of Medfly population relationships. They concluded that analysis of these alleles in ancestral African populations, where considerably more genetic variability occurs, is desirable.

13.7.3.4.3. Nuclear Gene Intron Size Variability

Gomulski et al. (1998) evaluated variability in the size of the first intron in the alcohol dehydrogenase gene to assess 16 populations from five geographical regions: Africa, the Mediterranean Basin, Latin America, Hawaii, and Australia. PCR primers were developed that spanned the first intron between exons 1 and 2. PCR product sizes varied from 1400 bp to 3450 bp and were grouped into four distinct categories: short, medium, long, and very long. Most variants were found only in the African populations and only a few migrated from Africa with the colonizing populations. The results obtained were congruent with those obtained by analyzing allozyme variation and showed a gradual and large reduction in intron variability. Gomulski et al. (1998) concluded that drift, bottleneck effects, and migration were important in explaining the observed intron size variability.

13.7.3.4.4. Multiple Nuclear Gene Intron Sequences

Multiple nuclear gene introns were analyzed by Villablanca et al. (1998) in an effort to provide sufficient information to resolve the origins of the Medfly populations in California

and Hawaii despite the expected reduction in variability due to founder effects and genetic bottlenecks. Villablanca et al. (1998) chose to work with multiple introns because they concluded that mitochondrial DNA is poorly suited to studies of invasions unless the invading population is large or grows rapidly: mtDNA is subject to strong genetic drift due to its maternal and haploid mode of inheritance. They also argued that RAPD-PCR data are difficult to interpret and may lack repeatability; microsatellites are effective, but require a long period of development for each new taxon. Their review of previous work of molecular analyses of Medflies indicated that "little genetic variation has been uncovered within invading populations with both allozymes . . . and mtDNA."

Using multiple intron sequences, Villablanca et al. (1998) "found a wealth of genetic variability within invading populations." Introns evolve more quickly than the protein-coding regions of a gene and are expected to retain variation due to their diploid and biparental inheritance. The intron sequence variation can be subjected to "phylogenetic analysis, cladistic analysis of gene flow, as well as standard population genetic and coalescence analysis of alleles" (Palumbi 1996, Roderick 1996). Medfly populations in Africa, California, Hawaii, Brazil, and Greece were evaluated. Primers were constructed, using Medfly sequence data from the literature, to amplify introns that have conserved positions across species from four single-copy nuclear genes. Single-copy genes were used in order to avoid analysis of nonspecific PCR products that could occur from multiple gene copies or pseudogenes. The four loci were: muscle-specific actin intron 1 (the other two actin loci known in Medfly are not muscle-specific), chorion s36 intron 1 (which is different from the other three chorion-like genes in Medfly), vitellogenin 1 gamma intron 2 (which is different from the other three known vitellogenins in Medfly), and Cu/Zn superoxide dismutase (SOD) intron 1 (a second Cu-Zn gene is known in Medfly, but has different amino acid sequences).

The PCR products were cloned and sequenced and were found to be specific to the particular targeted locus, with no evidence for pseudogenes. The sequence data were carefully analyzed to eliminate sequences in which errors were incorporated by *Taq* polymerase, which has a misincorporation rate of about one per 1000 bases. Villablanca et al. (1998) eliminated these erroneous sequences by sequencing between one and three clones per individual and then identifying and removing "singletons." Singletons are variability that occurs in only one sequence of an alignment. They occur in two forms: a particular nucleotide position is variable in only one sequence, or a unique sequence occurs in a variable position. "Not all singletons are PCR errors, but considering them to be so results in a conservative measure of allelic diversity" and the remaining sequences were analyzed phylogenetically by Templeton's network method, which allows reconstruction of phylogenies from potentially recombining DNA fragments (Villablanca et al. 1998).

Villablanca et al. (1998) interpreted the phylogenetic analysis of the four intron sequences as follows:

> The phylogeny of alleles shows that there is no phylogeographic structuring at the population level. Few alleles are shared between African and invading [California, Hawaii, Brazil, and Greece] populations. . . . The phylogenies of alleles, similarly, do not provide evidence that any invading population is monophyletic. . . . Although the phylogeny of alleles is not useful for phylogeographic analysis in this case, it is still essential in that it demonstrates that alleles might be shared among populations simply because all populations are ultimately derived from Africa and not because they share a common invasion history.

The authors pointed out that the next step is to "sample populations more thoroughly and test for population subdivision." Phylogeography is the study of relationships among

genotypes (phylogeny of alleles or haplotypes) from one or more populations that are examined relative to their geographical location (Roderick 1996).

Davies et al. (1999a) used the same multiple intron loci to distinguish between "alternative hypotheses concerning the source of medfly infestations in California." In this study, intron sequences from Villablanca et al. (1998) were used, as well as newly obtained intron sequences from Medfly samples in California, Costa Rica, Guatemala, Mexico, Brazil, Peru, Greece, Hawaii, and Africa. A total of 237 sequences were obtained for four loci in 74 individuals. The data from all Medflies in California were treated as a "single population for the purposes of statistical analysis. Under the null hypothesis that there is a resident medfly population in California, we assume that these flies, captured in the same geographic area, represent a single biological population" (Carey 1991).

To assess whether a recent outbreak was due to a new invasion, the authors focused on a single fly (B-96) captured in southern California (Burbank) in 1996. Analysis of molecular variation (AMOVA) produced indices of population subdivision analogous to standard F statistics. Another program, TFPGA, was used to calculate the average theta. Yet another program, IMMANC, was used to carry out an assignment test. Davies et al. (1999a) concluded, "Because the B-96 genotype was included in the Californian population (and not the potential source) when estimating the 'resident' allele frequencies, the test is conservative with respect to the null hypothesis that B-96 is a resident—in this case of California." Davies et al. (1999a) concluded that the single B-96 Medfly studied was "less likely ($\alpha < 0.05$) to be a resident of California than an immigrant from no less than four potential sources: Costa Rica, Guatemala, Mexico, or Peru." Finally, Davies et al. (1999a) concluded:

> More work is clearly needed to explore the phylogenetic consequences of invasions and a better understanding of invasion genetic patterns will provide a deeper insight into the ecological and evolutionary processes that underlie bioinvasions. It is important to consider . . . that invasions often involve a hierarchy of events, the totality of which might be termed a metainvasion. The metainvasion begins with a primary invasion, when a species first colonizes a new area from its ancestral source. Subsequently, secondary and tertiary invasions arise as the newly established populations themselves seed new areas. The genetic changes that result from these events are complex and phylogenetic analyses may be informative at some levels but not others. A primary invasion of the medfly occurred from Africa to the Mediterranean. The invasion of Latin America may be another primary invasion, direct from Africa, or a secondary invasion from the Mediterranean. Californian medfly invasions thus represent secondary or tertiary events in the global medfly metainvasion. Indeed, California may be subject to repeat invasions that could superimpose on one another.

Thus, by 1998 powerful genetic tools had been brought to bear on the Medfly colonization question, and the statistical methods for analyzing molecular data were becoming ever more sophisticated. Villablanca et al. (1998) suggested that the origin of Californian Medfly infestations might be determined through the use of microsatellites or single-strand conformation polymorphism analysis. They noted:

> Due to its economic significance, the medfly infestation has become a model system for the study of contemporary bioinvasions and has several important lessons for other cases where limited funding is likely to restrict the amount of research effort. The genetic analysis of new bioinvasions should begin with mtDNA and allozymes; however, highly variable nuclear regions, such as introns, should also be considered. Multilocus genotyping provides a rapid method of determining the origin of invasions, whether using nonsequencing methods of screening intron variation and/or other types of markers.

13.7.3.4.5. Microsatellite Analysis

Medfly microsatellites were proposed as potentially useful tools for population analysis by Bonizzoni et al. (2000), who screened a *C. capitata* genomic library to identify

30 microsatellite loci. In addition, 11 loci were identified by RAPD-PCR and random genomic sequencing. Two additional loci were identified in GenBank for a total of 43 microsatellites. Ten of these microsatellite sequences were used to analyze 122 Medflies from six populations (Africa: Kenya, Reunion, Madeira; Mediterranean: South Italy and Greece; and South America: Peru). The results obtained were "consistent with results obtained from allozyme and single-copy DNA studies with respect to the historically documented expansion of the medfly" (Bonizzoni et al. 2000). As with the allozyme data (Malacrida et al. 1998), polymorphisms decreased as flies moved from tropical Africa to the Mediterranean basin and to South America. Microsatellite analysis was extended by Bonizzoni et al. (2001) to analyze the invasion of Medflies into California: was there one established population or many invasive populations? The 10 previously characterized microsatellite loci were used to compare 109 Medflies captured in California between 1992 and 1998 with 242 Medflies from Hawaii, Guatemala, El Salvador, Ecuador, Brazil, Argentina, and Peru, using between six and 30 flies per sample site. Their data analysis used a method that accounts for heterogeneity in the size of samples to estimate allelic richness. The frequency of each allele per locus, the observed heterozygosity, and deviations from Hardy–Weinberg expectations were computed using several computer programs (Bonizzoni et al. 2001). Genetic divergence between individuals, as well as within and between populations, was estimated in terms of shared bands between individuals. Relationships between populations were given in dendrograms obtained from the dissimilarity index and Nei's unbiased genetic distance D_A. Trees were constructed using the neighbor-joining method of Felsenstein (1993); bootstrap values for the tree were obtained using the "gene frequency" option within the program SEQBOOT. The Kenyan sample was used as the outgroup because it is the most differentiated in the medfly species range. An estimation of the probability that the California Medflies are immigrants from South America, Central America or Hawaii was determined using the IMMANC test (Bonizzoni et al. 2001).

Bonizzoni et al. (2001) had three main conclusions:

> (i) Among the Latin American and Pacific samples, the Guatemalan flies are most closely related genetically to the California flies, according to the majority of the tests applied; (ii) the Californian infestations are structured, with the San Diego infestation being the most differentiated; and (iii) the fact that flies captured between 1992 and 1997 in the Los Angeles basin appear to be genetically related supports the hypothesis that an endemic population has been formed in this area.

Furthermore, "Hawaiian flies show relatively low similarities with Californian flies," which indicates Hawaii is not the source of the Californian Medflies tested, as was found previously by He and Haymer (1999).

Bonizzoni et al. (2001) further concluded that the situation could be more complex than expected:

> Within at least some of the Los Angeles basin samples, there is considerable evidence for genetic homogeneity. Based on this, the possibility of an endemic population in California cannot be excluded. It is entirely possible that independent infestations of this pest from the same geographical region, overlaid on an existing endemic population, have acted together to create this unique situation.

13.7.3.5. The End?

A California Department of Food and Agriculture press release in June 2001 summarized eradication costs since 1975 in California: "more than $256 million in state and federal

funds have been spent eradicating infestations of the pest, primarily in Southern California and the Bay Area." The report concluded that the 5-year test program to stop new infestations by releasing sterile Medflies in the Los Angeles basin was "enormously successful": "Between 1987 and 1994, an average of 7.5 Medfly infestations were discovered each year in California. Since the preventative Release Program began in 1996, there has been just one infestation for the entire five-year period."

Over the life of the 5-year "preventative program" that began in 1996, nearly 75 billion sterile Medflies were released by airplanes over more than 2000 square miles of the Los Angeles basin. This preventative program cost California approximately $7.4 million annually and the U.S. government an additional $7.4 million (CDFA Legislative Report, March 2000). The report highlighted the fact that "Since 1994, California's Medfly strategy in Southern California has shifted from a reactive approach, focused on detecting and eradicating early infestations, to a proactive approach that emphasizes preventative measures."

Stay tuned to learn whether the Medfly remains "eradicated" from southern California; whether the preventative sterile Medfly releases will be continued; whether established, low-density Medfly populations will rebound if the sterile Medfly releases are discontinued in the Los Angles basin; whether new invasions are detected; and whether further resolution of the number of invasions and their sources can be obtained. Will the population genetic structure of Medfly populations in California become too complex to unravel if additional invasions occur?

13.7.3.6. Some Lessons Learned

This case study illustrates several points: molecular tools vary in their sensitivity, ease of use, cost, and time to develop. Despite differences in methods used, the various researchers agreed on some key findings: Medfly populations in the Los Angeles area were unlikely to have come from Hawaii; the population present in the Los Angeles area may be the result of an established population, multiple infestations from the same source, or both.

Molecular markers of six Medflies from the San Diego population were clearly different from the markers found in Medflies collected in the Los Angeles area. The San Diego Medflies sampled most closely resembled Medflies from Hawaii. Analysis of the infestation area by entomologists correlated the San Diego infestation with a family that had just returned from a trip to Hawaii. Thus, combining molecular and ecological data may provide more information than relying on a single approach (A. Malacrida, personal communication).

It should be clear that insect population genetic structure can be quite complex; simply assuming that molecular markers will allow an unequivocal conclusion as to the population's geographic origin may be unrealistic, especially if primary, secondary, and tertiary invasions have occurred (Figure 13.4). Genetic variability and structuring is dependent upon events in the population's history, including bottlenecks, drift, selection and hybridization. The Medfly case study shows that increasingly refined molecular methods are available, and the statistical and other analyses used to reach conclusions are becoming more sophisticated.

13.7.4. Plant Defenses to Insect Herbivory

Plants face a variety of biotic (bacteria, fungi, insects, nematodes and other herbivores) and abiotic (drought, heat, salinity, UV damage) stresses in their environment (for a review, see Strauss and Agrawal 1999). In response, plants evolved both constitutive and inducible defenses that have a genetic basis. There appear to be many genes involved in

plant defense mechanisms, making it difficult to analyze their role using traditional genetic methods.

DNA microarrays were used to obtain significant advances in our understanding of how plants defend themselves against insect herbivores (Schenk et al. 2000, Baldwin et al. 2001, Reymond 2001). An analysis of approximately 7000 *Arabidopsis* genes (which is 25 to 30% of this plant's genome) suggests that about 300 of the 7000 genes (4.3%) evaluated are involved in defense (Maleck et al. 2000). Comparison of gene transcript profiles after plants were exposed to different stresses revealed that some genes have overlapping roles in defense; thus, microarray analyses can elucidate how plants respond to multiple stressors.

Microarrays of *Arabidopsis* expressed sequence tags (ESTs) have become available from several core laboratory sources since this species' genome was sequenced (Reymond 2001). The microarrays allow simultaneous hybridization of probes to an array of immobilized DNA fragments that correspond to a specific gene. After scanning the microarray with a laser scanner, the signal for each fragment reflects the abundance of the corresponding messenger RNA in the sample (Maleck et al. 2000).

Many plants produce volatile organic compounds (VOCs) in response to feeding damage caused by herbivores such as insects (Dicke 1999). These VOCs may influence neighboring plants to respond rapidly to wounding, or to insect or plant pathogen attack. Some VOCs attract predators and parasitoids to the insect herbivore, and DNA microarrays will allow researchers to analyze which genes are involved in this type of plant defense. DNA microarrays will allow researchers "to determine the extent to which VOCs can elicit defense-related transcripts in neighboring plants" (Arimura et al. 2000a,b).

It appears that a complex network of interdependent signaling pathways convey molecular messages in *Arabidopsis* that identify the type of pest, which allows the plant to mount an appropriate response (Reymond 2001). Integrating the information on plant responses, gaining an understanding of the communications that take place between the different defense response pathways, and obtaining a complete list of response genes should be achievable. Thus, the global analysis of plant gene expression in microarrays and the complete sequencing of the *Arabidopsis* and rice genomes should revolutionize the analysis of insect–plant interactions in the near future. Microarray research will bring together ecologists, molecular biologists, and plant scientists (Maleck et al. 2000, Baldwin et al. 2001). Some caution is needed, however, because conducting and analyzing microarray experiments requires careful consideration of experimental design and statistical analysis (Kerr and Churchill 2001a,b, Quackenbush 2001).

13.8. Transgenic Organisms and Evolutionary Ecology?

Tatar (2000) noted that evolutionary ecologists attempt to understand how particular populations or individuals survive and reproduce over time. An experimental approach to this goal would to be to alter a trait genetically and observe the resulting change in reproduction, survival, growth, defense, or competitive ability. Tatar (2000) noted that, with the use of transgenic technology, it is possible to introduce novel genes into the germ line and have them be expressed, which could provide evolutionary ecologists with a powerful tool to understand the mechanisms underlying adaptive traits and their evolution.

For example, research on transgenic *D. melanogaster* indicates that some genes may affect longevity (Tatar et al. 2001). Overexpression of heat shock protein 70 was

evaluated in *D. melanogaster* as a mechanism for increasing longevity and stress response. Unfortunately, extra copies of heat shock protein 70 may not be beneficial under all circumstances (Feder et al. 1997, Feder 1999, Silbermann and Tatar 2000). A number of studies report deleterious effects of large amounts of heat shock proteins in *D. melanogaster*, perhaps because the proteins consume so much energy and substrate that other important functions are affected negatively (Feder 1999).

The release of transgenic insects into the environment cannot be conducted without obtaining permission from several regulatory authorities and after risk analyses (see Chapter 14 for a discussion). Thus, a transgenic approach to analyzing evolutionary ecology questions may be limited to laboratory or contained field plots.

13.9. Applied Pest Management

Intraspecific variation and genetic change in both pest and beneficial arthropods influence pest management strategies and tactics in different ways. The fact that pest and natural enemy populations change genetically has been discussed often. Molecular genetic techniques allow for rapid assessment of intraspecific genetic variation, changes in host preferences, insecticide resistance levels, and pest and natural enemy biotypes significant in biological control of pest arthropods and weeds. However, molecular methods are not a panacea, and Curtis (2002) recently raised a relevant question.

13.9.1. The “So What?” Test

Curtis (2002) reviewed three areas in medical entomology where molecular methods have been applied. These include identifying complexes of sibling malaria vector species, evaluating insecticide resistance in vectors of malaria, and developing nonsusceptibility to pathogens in mosquitoes using transgenic methods. Curtis (2002) questioned the number of instances in which “molecular taxonomic markers for these characteristics add to what can be measured simply and directly” and noted “There are very few instances where control programmes are guided by data on sibling species compositions of vector populations.”

Likewise, Curtis (2002) noted, “Much emphasis is placed on studies of biochemical or molecular resistance mechanisms but it was painstaking and direct field testing, rather than these molecular or biochemical studies, that revealed” resistance to pyrethroid insecticides in malaria vectors in South Africa. Curtis (2002) concluded

> that molecular methods in medical entomology should not be an end in themselves. They should be adopted only after careful investigation shows that they can pass the “so what?” test, that is, could they solve real problems that are not already solvable by simpler and cheaper methods.

There are genuine benefits to molecular methods, although they are not panaceas. Because the field is yet young, it is difficult to know *a priori* whether a particular molecular tool will aid in solving a specific problem. However, it is appropriate to ask whether there are more rapid and less expensive methods that can resolve the question.

13.10. Relevant Journals

Biochemical Systematics and Ecology, Pergamon Press
Conservation Genetics, Kluwer Academic Publ., The Netherlands

Molecular Ecology, Blackwell Scientific Publ., Oxford

Molecular Ecology Notes, Blackwell Scientific Publ., Oxford. This is a sister journal to *Molecular Ecology* and will contain primer, computer, and technical notes; papers are published on line, as well as in print.

References Cited

Abate, T. (1993). Is the invading medfly eligible for California residency? *BioScience* **43**: 3–6.

Agusti, N. M. C. De Vicente, and R. Gabarra (1999). Development of sequence characterized region (SCAR) markers of *Helicoverpa armigera*: a new polymerase chain reaction-based technique for predator gut analysis. *Mol. Ecol.* **8**: 1476–1474.

Aquadro, C. F., W. A. Noon, and D. J. Begun (1992). RFLP analysis using heterologous probes, pp. 115–158. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

Arimura, G., K. Tashiro, S. Kuhara, T. Nishioka, R. Ozawa, and J. Takabayashi (2000a). Gene responses in bean leaves induced by herbivory and by herbivore-induced volatiles. *Biochem. Biophys. Res. Commun.* **277**: 305–310.

Arimura, G., R. Ozawa, T. Shimoda, T. Nishioka, W. Boland, and J. Takabyashi (2000b). Herbivory-induced volatiles elicit defence genes in lima bean leaves. *Nature* **406**: 512–515.

Armstrong, K. F., C. M. Cameron, and E. R. Frampton (1997). Fruit fly (Diptera: Tephritidae) species identification: a rapid molecular diagnostic technique for quarantine application. *Bull. Entomol. Res.* **87**: 111–118.

Arnheim, N., T. White, and W. E. Rainey (1990). Application of PCR: organismal and population biology. *BioScience* **40**: 174–182.

Ashley, M. V., and B. D. Dow (1994). The use of microsatellite analysis in population biology: background, methods and potential applications, pp. 187–201. In: *Molecular Ecology and Evolution Approaches and Applications*. B. Schierwaer, B. Stret, G. P. Wagner, and R. DeSalle, Eds. Birkhauser Verlag, Basel, Switzerland.

Atkinson, L., and E. S. Adams (1997). Double-strand conformation polymorphism (DSCP) analysis of the mitochondrial control region generates highly variable markers for population studies in a social insect. *Insect Mol. Biol.* **6**: 369–376.

Attimonelli, M., N. Altamura, R. Benne, A. Brennicke, J. M. Cooper, D. D'Ellia, A. de Montalvo, et al. (2000). MitBASE: a comprehensive and integrated mitochondrial DNA database. The present status. *Nucleic Acids Res.* **28**: 148–152.

Baldwin, I. T., R. Halitschke, A. Kessler, and U. Schittko (2001). Merging molecular and ecological approaches in plant-insect interactions. *Curr. Opin. Plant Biol.* **4**: 351–358.

Balloux, F., and N. Lugon-Moulin (2002). The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* **11**: 155–165.

Baruffi, l. G. Damiani, C. R. Guglielmino, C. Bandi, A. R. Malacrida, and G. Gasperi (1995). Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data. *Heredity* **74**: 425–537.

Berticat, C., M. P. Dubois, M. Marquine, C. Chevillon, and M. Raymond (2000). A molecular test to identify resistance alleles at the amplified esterase locus in the mosquito *Culex pipiens*. *Pest Management Sci.* **56**: 727–731.

Black, W. C., IV, N. M. DuTeau, G. J. Puterka, J. R. Nechols, and J. M. Pettorini (1992). Use of the random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) to detect DNA polymorphisms in aphids (Homoptera: Aphididae). *Bull. Entomol. Res.* **82**: 151–159.

Black, W. C. IV, C. F. Baer, M. F. Antolin, and N. M. DuTeau (2001). Population genomics: genome-wide sampling of insect populations. *Annu. Rev. Entomol.* **46**: 441–469.

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *Quart. Rev. Biol.* **74**: 21–45.

Bonizzoni, M., A. R. Malacrida, C. R. Guglielmino, L. M. Gomulski, G. Gasperi, and L. Zheng (2000). Microsatellite polymorphism in the Mediterranean fruit fly, *Ceratitis capitata*. *Insect Mol. Biol.* **9**: 251–261.

Bonizzoni, M., L. Zheng, C. R. Guglielmino, D. S. Haymer, G. Gasperi, L. M. Gomulski, and A. R. Malacrida (2001). Microsatellite analysis of medfly bioinfestations in California. *Mol. Ecol.* **10**: 2525–2524.

Bossart, J. L., and D. Pashley Prowell (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol. Evol.* **13**: 202–206.

Brookfield, J. F. Y. (1992). DNA fingerprinting in clonal organisms. *Mol. Ecol.* **1**: 21–26.

Bruford, M. W. O. Hanotte, J. F. Y. Brookfield, and T. Burke (1992). Single-locus and multilocus DNA fingerprinting, pp. 225–269. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

Buckley, T. R., C. Simon, P. K. Flook, and B. Misof (2000). Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. *Insect Mol. Biol.* **9**: 565–580.

Burke, T. (1989). DNA fingerprinting and other methods for the study of mating success. *Trends Ecol. Evol.* **4**: 139–144.

California Department of Food and Agriculture Report to the Legislature. (March 2000.) Mediterranean Fruit Fly Preventative Release Program.

California Department of Food and Agriculture (2002.) The Mediterranean Fruit Fly Fact Sheet. http://134.186.235.98/pests/medfly/mediterranean_fly.html

Carey, J. R. (1991). Establishment of the Mediterranean fruit fly in California. *Science* **253**: 1369–1373.

Carey, J. R. (1996a). The incipient Mediterranean fruit fly population in California Implications for invasion biology? *Ecology* **77**: 1690–1697.

Carey, J. R. (1996b). The future of the Mediterranean fruit fly *Ceratitis capitata* invasion of California: a predictive framework. *Biol. Conserv.* **78**: 35–50.

Cavalli-Sforza, L. L. (1998). The DNA revolution in population genetics. *Trends Genet.* **14**: 60–65.

Cooley, J. R., C. Simon, D. C. Marshall, K. Slon, and C. Ehrhardt (2001). Allochronic speciation, secondary contact, and reproductive character displacement in periodical cicadas (Hemiptera: *Magicicada* spp.): genetic, morphological and behavioural evidence. *Mol. Ecol.* **10**: 661–671.

Cornuet, J. M., S. Piry, G. Luikart, A. Estoup, and M. Solignac (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**: 1989–2000.

Crozier, R. H. (1977). Evolutionary genetics of the Hymenoptera. *Annu. Rev. Entomol.* **22**: 263–288.

Cullings, K. W., and D. R. Vogler (1998). A 5.8S nuclear ribosomal RNA gene sequence database: applications to ecology and evolution. *Mol. Ecol.* **7**: 919–923.

Curtis, C. (2002). Molecular medical entomology and the "so what?" test. *Trends Ecol. Evol.* **17**: 102.

Davies, N., F. X. Villablanca, and G. K. Roderick (1999a). Bioinvasions of the Medfly *Ceratitis capitata*: source estimation using DNA sequences at multiple intron loci. *Genetics* **153**: 351–360.

Davies, N., F. X. Villablanca, and G. K. Roderick (1999b). Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends Ecol. Evol.* **14**: 17–21.

Dick, M., D. M. Bridge, W. C. Wheeler, and R. DeSalle (1993). Collection and storage of invertebrate samples, pp. 51–65. In: *Methods in Enzymology.* Vol. (224). E. A. Zimmer, Ed., Academic Press, San Diego, CA.

Dicke, M. (1999). Evolution of induced indirect defence of plants, pp. 62–88. In: *The Ecology and Evolution of Inducible Defences*. R. Tollrian, and C. D. Harvell, Eds. Princeton Univ. Press, Princeton, NJ.

Doolittle, R. F., Ed. (1990). *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Vol. (183), *Methods in Enzymology*. Academic Press, New York.

Douglas, L. J., and D. S. Haymer (2001). Ribosomal ITS1 polymorphisms in *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae). *Ann. Entomol. Soc. Am.* **94**: 726–731.

Dowling, T. E., C. Moritz, and J. D. Palmer (1990). Nucleic acids II: restriction site analysis, pp. 250–317. In: *Molecular Systematics*. D. M. Hillis, and C. Moritz, Eds. Sinauer Assoc., Sunderland, MA.

Edwards, O. W., and M. A. Hoy (1993). Polymorphism in two parasitoids detected using random amplified polymorphic DNA (RAPD) PCR. *Biol. Control* **3**: 243–257.

Edwards, O. W., and M. A. Hoy (1995). Random amplified polymorphic DNA markers to monitor laboratory-selected, pesticide-resistant *Trioxys pallidus* (Hymenoptera: Aphidiidae) after release into three California walnut orchards. *Environ. Entomol.* **24**: 487–496.

Erlich, H. A., Ed. (1989). *PCR Technology, Principles and Applications of DNA Amplification*, Stockton Press, New York.

Estoup, A., and B. Angers (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations, pp. 55–86. In: *Advances in Molecular Ecology*. G. R. Carvalho, Ed. IOS Press, Amsterdam.

Estoup, A. M. Solignac, M. Harry, and J. M. Cornuet (1993). Characterization of $(GT)_n$ and $(CT)_n$ microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Res.* **21**: 1427–1431.

Feder, M. E. (1999). Organismal, ecological, and evolutionary aspects of heat-shock proteins and the stress response: established conclusions and unresolved issues. *Am. Zool.* **39**: 857–864.

Feder, M. E. N. V. Cartano, L. Milos, R. A. Krebs, and S. L. Lindquist (1997). Effect of engineering *Hsp70* copy number on Hsp70 expression and tolerance of ecologically relevant heat shock in larvae and pupae of *Drosophila melanogaster*. *J. Exp. Biol.* **119**: 1637–1844.

Feder, J. L., C. A. Chilcote, and G. L. Bush (1988). Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature* **336**: 61–64.

Felsenstein, J. (1993). *Phylip* (*Phylogeny Inference Package*). v 3.Sc. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Ferraris, J. D., and S. R. Palumbi (1996). *Molecular Zoology. Advances, Strategies, and Protocols.* Wiley-Liss, New York.

ffrench-Constant, R. H., N. Anthony, K. Aronstein, T. Rocheleau, and G. Stilwell (2000). Cyclodiene insecticide resistance: from molecular to population genetics. *Annu. Rev. Entomol*. **48**: 449–466.

FLYBase Consortium (2002). The FLYBase database of the *Drosophila* genome projects and community literature. *Nucleic Acids Res.* **30**: 106–108.

Gibson, G. (2002). Microarrays in ecology and evolution: a preview. *Mol. Ecol.* **11**: 17–24.

Gibson, G., and S. V. Muse (2002). *A Primer of Genome Science*. Sinauer Assoc., Sunderland, MA.

Gilbert, D. G. (2002). euGenes: a eukaryote genome information system. *Nucleic Acids Res.* **30**: 145–148.

Gomulski, L. M., K. Bourtzis, S. Brogna, P. A. Morandi, C. Bonvicini, F. Sebastiani, C. Torti, C. R. Guglielmino, C. Savakis, G. Gasperi, and A. R. Malacrida (1998). Intron size polymorphism of the Adh(1) gene parallels the worldwide colonization history of the Mediterranean fruit fly, *Ceratitis capitata*. *Mol. Ecol.* **7**: 1729–1741.

Goodman, S. J. (1997). R_{ST} Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Mol. Ecol.* **6**: 881–885.

Goodnight, K. F., and D. C. Queller (1999). Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Mol. Ecol.* **8**: 1231–1234.

Gribskov, M., and J. Devereux, Eds. (1991). *Sequence Analysis Primer*. Stockton Press, New York.

Hadrys, H., M. Balick, and B. Schierwater (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol. Ecol.* **1**: 55–63.

Hadrys, H., B. Schierwater, S. L. Dellaporta, R. DeSalle, and L. W. Buss (1993). Determination of paternity in dragonflies by random amplified polymorphic DNA fingerprinting. *Mol. Ecol.* **2**: 79–87.

Hagerman, P. J. (1990). Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* **59**: 755–781.

Hartl, D. L. (1981). *A Primer of Population Genetics*. Sinauer Assoc., Sunderland, MA.

Hartl, D. L., and A. G. Clark (1997). *Principles of Population Genetics*. 3rd ed. Sinauer, Sunderland, MA.

Haymer, D. S. (1994). Arbitrary (RAPD) primer sequences used in insect studies. *Insect Mol. Biol.* **3**: 191–194.

Haymer, D. S., and D. O. McInnis (1994). Resolution of populations of the Mediterranean fruit fly at the DNA level using random primers for the polymerase chain reaction. *Genome* **37**: 244–248.

Haymer, D. S., T. Tanaka, and C. Teramae (1994). DNA probes can be used to discriminate between tephritid species at all stages of the life cycle (Diptera: Tephritidae). *J. Econ. Entomol.* **87**: 741–746.

Headrick, D. H., and R. D. Goeden (1996). Issues concerning the eradication or establishment and biological control of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae), in California. *Biol. Control* **6**: 412–421.

He, M., and D. S. Haymer (1999). Genetic relationships of populations and the origins of new infestations of the Mediterranean fruit fly. *Mol. Ecol.* **8**: 1247–1257.

Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Mol. Ecol.* **10**: 537–549.

Hickson, R. E., C. Simon, A. Cooper, G. S. Spicer, J. Sullivan, and D. Penny (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Mol. Biol. Evol.* **13**: 150–160.

Hoelzel, A. R., and D. R. Bancroft (1992). Statistical analysis of variation, pp. 297–305. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

Hoelzel, A. R., and G. A. Dover (1991). Statistical interpretation of variation and genetic distance, pp. 31–46. In: *Molecular Genetic Ecology.* IRL Press, Oxford.

Hoelzel, A. R., and A. Green (1992). Analysis of population-level variation by sequencing PCR-amplified DNA, pp. 159–187. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford.

Hoffmann, A. A., C. M. Sgro, and S. H. Lawler (1995). Ecological population genetics: the interface between genes and the environment. *Annu. Rev. Genet.* **29**: 349–370.

Hoogendorn, M., and G. E. Heimpel (2001). PCR-based gut content analysis of insect predators: using ribosomal ITS-1 fragments from prey to estimate predation frequency. *Mol. Ecol.* **10**: 2059–2067.

Howard, D. J., and S. H. Berlocher, Eds. (1998). *Endless Forms: Species and Speciation.* Oxford Univ. Press, New York.

Hoy, M. A., A. Jeyaprakash, and R. Nguyen (2001). Long PCR is a sensitive method for detecting *Liberobacter asiaticum* in parasitoids undergoing risk assessment in quarantine. *Biol. Control* **22**: 278–287.

Huffaker, C. B., and A. P. Gutierrez, Eds. (1999). *Ecological Entomology*, 2nd ed. J. Wiley & Sons, New York.

Innis, M. A., D. H. Gelfand, J. J. Ninsky, and T. J. White, Eds. (1990). *PCR Protocols, A Guide to Methods and Applications*. Academic Press, San Diego.

Jeffreys, A. J. (1987). Highly variable minisatellites and DNA fingerprints. *Biochem. Soc. Trans.* **15**: 309–317.

Jeyaprakash, A., and M. A. Hoy (2000). Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of 63 arthropod species. *Insect Mol. Biol.* **9**: 393–405.

Kakouli-Duarte, T., D. G. Casey, and A. M. Burnell (2001). Development of a diagnostic DNA probe for the fruit flies *Ceratitis capitata* and *Ceratitis rosa* (Diptera: Tephritidae) using amplified fragment-length polymorphism. *J. Econ. Entomol.* **94**: 989–997.

Kambhampati, S., W. C. Black IV, and K. S. Rai (1991). Geographic origin of the US and Brazilian *Aedes albopictus* inferred from allozyme analysis. *Heredity* **67**: 85–94.

Kambhampati, S., W. C. Black, and K. S. Rai (1992). A RAPD-PCR based method for the identification and differentiation of mosquito populations: techniques and statistical analyses. *J. Med. Entomol.* **29**: 939–945.

Karl, S. A., and J. C. Avise (1993). PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. *Mol. Biol. Evol.* **10**: 342–361.

Kashi, Y., D. King, and M. Soller (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* **13**: 74–78.

Kellenberger, E. (1994). Genetic ecology: a new interdisciplinary science, fundamental for evolution, biodiversity and biosafety evaluations. *Experientia* **50**: 429–437.

Kerr, M. K., and G. A. Churchill (2001a). Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. USA* **98**: 8961–8965.

Kerr, M. K., and G. A. Churchill (2001b). Experimental design for gene expression microarrays. *Biostatistics* **2**: 183–201.

Kirby, L. T. (1990). *DNA Fingerprinting. An Introduction.* Stockton Press, New York.

Lambert, D. M. (1995). The new science of molecular ecology. N. Z. *J. Ecol.* **19**: 93–96.

Lanave, C., F. Licciulli, M. De Robertis, A. Marolla, and M. Attimonelli (2002). Update of AMmtDB: a database of multi-aligned Metazoa mitochondrial DNA sequences. *Nucleic Acids Res.* **30**: 174–175.

Landry, B. S., L. Dextraze, and G. Boivin (1993). Random amplified polymorphic DNA markers for DNA fingerprinting and genetic variability assessment of minute parasitic wasp species (Hymenoptera: Mymaridae and Trichogrammatidae) used in biological control programs of phytophagous insects. *Genome* **36**: 580–587.

Lester, L. J., and R. K. Selander (1979). Population genetics of haplodiploid insects. *Genetics* **92**: 1329–1345.

Loxdale, H. D., and G. Luschai (1998). Molecular markers in entomology. *Bull. Entomol. Res.* **88**: 577–600.

Lu, Y. J., M. J. Adang, D. J. Isenhour, and G. D. Kochert (1992). RFLP analysis of genetic variation in North American populations of the fall armyworm moth *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Mol. Ecol.* **1**: 199–208.

Lu, Y., G. D. Kochert, D. J. Eisenhour, and M. J. Adang (1994). Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Insect Mol. Biol.* **3**: 123–130.

Luikart, G., and P. R. England (1999). Statistical analysis of microsatellite DNA data. *Trends Ecol. Evol.* **14**: 253–256.

Lunt, D. H., L. E. Whipple, and B. C. Hyman (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol. Ecol.* **7**: 1441–1455.

Lynch, M., and T. J. Crease (1990). The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.* **7**: 377–394.

MacPherson, J. M., P. E. Eckstein, G. J. Scoles, and A. A. Gajadhar (1993). Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Mol. Cell. Probes* **7**: 293–299.

Malacrida, A. R., C. R. Guglielmino, P. D. Adamo, C. Torti, F. Marinoni, and G. Gasperi (1996). Allozyme divergence and phylogenetic relationships among species of tephritid flies. *Heredity* **76**: 592–602.

Malacrida, A. R., F. Marinoni, C. Torti, L. M. Gomulski, F. Sebastiani, C. Vonviccini, G. Gasperi, and C. R. Guglielmino (1998). Genetic aspects of the worldwide colonization process of *Ceratitis capitata*. *Genetics* **89**: 501–507.

Maleck, K., A. Levine, T. Eulgem, A. Morgan, J. Schmid, K. A. Lawton, J. L. Dangl, and R. A. Dietrich (2000). The transcriptome of *Arabidopsis* during systemic acquired resistance. *Nat. Genet.* **26**: 403–410.

Marshall, D. C. (2001). Periodical cicada (Homoptera: Cicadidae) life-cycle variations, the historical emergence record, and the geographic stability of brood distributions. *Ann. Entomol. Soc. Am.* **94**: 386–399.

Marshall, D. C., and J. R. Cooley (2000). Reproductive character displacement and speciation in periodical cicadas, with description of a new species, 13-year *Magicicada neotredecim*. *Evolution* **54**: 1313–1325.

Martin, A. P., and C. Simon (1988). Anomalous distribution of nuclear and mitochondrial DNA markers in periodical cicadas. *Nature* **336**: 6196.

May, B. (1992). Starch gel electrophoresis of allozymes, pp. 1–27. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

McMichael, M., and D. Pashley Prowell (1999). Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strains. *Ann. Entomol. Soc. Am.* **92**: 175–181.

Menken, S. B. J. (1991). Does haplodiploidy explain reduced levels of genetic variability in Hymenoptera? *Proc. Exp. Appl. Entomol. N.E.V. Amsterdam* **2**: 172–178.

Mitton, J. B. (1994). Molecular approaches to population biology. *Annu. Rev. Ecol. Syst.* **25**: 45–69.

Moritz, C., T. E. Dowling, and W. M. Brown (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* **18**: 269–192.

Mount, D. W. (2001). *Bioinformatics. Sequence and Genome Analysis.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mueller, U. G., and L. L. Wolfenbarger (1999). AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* **14**: 389–393.

Murphy, R. W., J. W. Sites, Jr., D. G. Buth, and C. H. Haufler (1990). Proteins I: Isozyme electrophoresis, pp. 45–126. In: *Molecular Systematics*. D. M. Hillis, and C. Moritz, Eds. Sinauer Assoc., Sunderland, MA.

Myers, J. H., D. Simberloff, A. M. Kuris, and J. R. Carey (2000). Eradication revisited: dealing with exotic species. *Trends Ecol. Evol.* **15**: 316–320.

Nei, M. (1972). Genetic distance between populations. *Am. Natur.* **106**: 283–292.

Neve, G., and E. Meglecz (2000). Microsatellite frequencies in different taxa. *Trends Ecol. Evol.* **15**: 376–377.

Odum, E. P. (1971). *Fundamentals of Ecology*. 3rd ed. W. B. Saunders, Philadelphia, PA.

Packer, L., and R. E. Owen (1992). Variable enzyme systems in the Hymenoptera. *Biochem. Syst. Ecol.* **20**: 1–7.

Palumbi, S. R. (1996). Macrospatial genetic structure and speciation in taxa with high dispersal abilities, pp. 101–108. In: *Molecular Zoology: Advances, Strategies, and Protocols.* J. Ferraris, and S. R. Palumbi, Eds. Wiley, New York.

Pashley, D. P. (1986). Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae): a sibling species complex? *Ann. Entomol. Soc. Amer.* **79**: 898–904.

Pashley, D. P. (1988). Quantitative genetics, development, and physiological adaptation in host strains of fall armyworm. *Evolution* **42**: 93–102.

Pashley, D. P. (1989). Host-associated differentiation in armyworms (Lepidoptera: Noctuidae): an allozymic and mitochondrial DNA perspective, pp. 103–114. In: *Electrophoretic Studies on Agricultural Pests*. H. D. Loscale, and J. den Hollander, Eds. Clarendon Press, Oxford, UK.

Pasteur, N., G. Pasteur, F. Bonhomme, J. Catalan, and J. Britton-Davidian (1988). *Practical Isozyme Genetics*. Ellis Horwood, Chichester, UK.

Penner, G. A., A. Bush, R. Wise, W. Kim, L. Domier, K. Kasha, A. Laroche, G. Scoles, S. J. Molar, and G. Fedak (1993). Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *PCR Methods Appl.* **2**: 341–345.

Price, P. W. (1997). *Insect Ecology.* 3rd ed. John Wiley, New York.

Quackenbush, J. (2001). Computational analysis of microarray data. *Nat. Rev. Genet.* **2**: 418–427.

Real, L. A., Ed. (1994). *Ecological Genetics*. Princeton Univ. Press, Princeton, NJ.

Reymond, P. (2001). DNA microarrays and plant defense. *Plant. Physiol. Biochem.* **39**: 313–321.

Rieseberg, L. H. (1998). Molecular ecology of hybridization, pp. 243–265. In: *Advances in Molecular Ecology*, G. R. Carvalho, Ed. IOS Press, Amsterdam.

Roderick, G. K. (1996). Geographic structure of insect populations: gene flow, phylogeography and their uses. *Annu. Rev. Entomol.* **41**: 263–290.

Roehrdanz, R. L., and M. E. DeGrugillier (1998). Long sections of mitochondrial DNA amplified from fourteen orders of insects using conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **91**: 771–778.

Rozas, J., and R. Rozas (1997). DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *CABIOS* **13**: 307–311.

Saad, F. A., B. Halliger, C. R. Muller, R. G. Roberts, and G. A. Danieli (1994). Single base substitutions are detected by double strand conformation analysis. *Nucleic Acids Res.* **22**: 4352–4353.

Schenk, P. M., K. Kazan, I. Wilson, J. P. Anderson, T. Richmond, S. C. Somerville, and J. M. Manners (2000). Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **97**: 11655–11660.

Schnabel, A., P. Beerli, A. Estoup, and D. Hillis (1998). A guide to software packages for data analysis in molecular ecology, pp. 291–303. In: *Advances in Molecular Ecology*. G. R. Carvalho, Ed. IOS Press, Amsterdam.

Schwartz, M. K., D. A. Tallmon, and G. Luikart (1998). Review of DNA-based census and effective population size estimators. *Animal Conserv.* **1**: 293–299.

Silbermann, R., and M. Tatar (2000). Reproductive costs of heat shock protein in transgenic *Drosophila melanogaster*. *Evolution* **54**: 2038–2045.

Simon, C., A. Franke, and A. Martin (1991). The polymerase chain reaction: DNA extraction and amplification, pp. 329–355. In: *Molecular Techniques in Taxonomy*. G. M. Hewitt, A. W. B. Johnston, and J. P. W. Young, Eds. Springer-Verlag, Berlin.

Simon, C., C. McIntosh, and J. Deniega (1993). Standard restriction fragment length analysis of the mitochondrial genome is not sensitive enough for phylogenetic analysis or identification of 17-year periodical cicada broods (Hemiptera: Cicadidae): the potential for a new technique. *Ann. Entomol. Soc. Am.* **86**: 228–238.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook (1994). Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**: 651–701.

Simon, C. J. Tang, S. Dalwadi, G. Staley, J. Deniega, and T. R. Unnasch (2000). Genetic evidence for assortative mating between 13-year cicadas and sympatric "17-year cicadas with 13-year life cycles" provides support for allochronic speciation. *Evolution* **54**: 1326–1336.

Slatkin, M. (1987). Gene flow and the geographic structure of a natural population. *Science* **236**: 787–792.

Slatkin, M., and N. H. Barton (1989). A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**: 1349–1368.

Sonvico, A., F. Manso, L. A. Quesada Allue (1996). Discrimination between the immature stages of *Ceratitis capitata* and *Anastrepha fraterculus* (Diptera: Tephritidae) populations by random amplified polymorphic DNA polymerase chain reaction. *J. Econ. Entomol.* **89**: 1208–1212.

Southwood, T. R. E. (1978). *Ecological Methods*. 2nd ed. Halsted, New York.

Spanos, L., G. Koutroumbas, M. Kotsyfakis, and C. Louis (2000). The mitochondrial genome of the Mediterranean fruit fly. *Insect Mol. Biol.* **9**: 139–144.

Strauss, S. Y., and A. A. Agrawal (1999). The ecology and evolution of plant tolerance to herbivory. *Trends Ecol. Evol.* **14**: 179–185.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends Ecol. Evol.* **15**: 199–203.

Symondson, W. O. C., and J. E. Liddell, Eds. (1996). *The Ecology of Agricultural Pests. Biochemical Approaches.* Chapman & Hall, London, UK.

Szymanski, M., M. Z. Barciszewska, V. A. Erdmann, and J. Barcisewski (2002). 5S ribosomal RNA database. *Nucleic Acids Res.* **30**: 176–178.

Tabachnick, W. J., and W. C. Black IV (1995). Making a case for molecular population genetic studies of arthropod vectors. *Parasitol. Today* **11**: 27–29.

Tang, J., and T. R. Unnasch (1997). Heteroduplex analysis in medical entomology: a rapid and sensitive sequence-based tool for population and phylogenetic studies. *Parasitol. Today* **13**: 271–274.

Tatar, M. (2000). Transgenic organisms in evolutionary ecology. *Trends Ecol. Evol.* **15**: 207–211.

Tatar, M., A. Kopelman, D. Epstein, M. P. Tu, C. M. Yin, and R. S. Garofalo (2001). A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**: 107–110.

Taylor, M. F. J., S. W. McKechnie, N. Pierce, and M. Kreitman (1993). The Lepidopteran mitochondrial control region: structure and evolution. *Mol. Biol. Evol.* **10**: 1259–1272.

Tegelstrom, H. (1992). Detection of mitochondrial DNA fragments. pp. 89–113. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

Templeton, A. R. (1998). Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Mol. Ecol.* **7**: 381–397.

Toth, G., Z. Gaspari, and J. Jurka (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**: 967–981.

U.S. Department of Agriculture, Animal and Plant Health Inspection Service (2002). The Mediterranean fruit fly. http://www.aphis.usda.gov/oa/pubs/fsmedfly.html

Villablanca, F. X., G. K. Roderick, and S. R. Palumbi (1998). Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Mol. Ecol.* **7**: 547–560.

Wang, J., and A. Caballero (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**: 212–226.

Wang, R., F. C. Kafatos, and L. Zheng (1999). Microsatellite markers and genotyping procedures for *Anopheles gambiae*. *Parasitol. Today* **15**: 33–37.

Weir, B. S. (1990). *Genetic Data Analysis. Methods for Discrete Population Genetic Data*. Sinauer Assoc., Sunderland, MA.

Weir, B. S., and C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Weiss, P. (1950). Perspectives in the field of morphogenesis. *Quart. Rev. Biol.* **25**: 177–198.

White, P. S., and L. D. Densmore III. (1992). Mitochondrial DNA isolation, pp. 29–58. In: *Molecular Genetic Analysis of Populations. A Practical Approach.* A. R. Hoelzel, Ed. IRL Press, Oxford, UK.

Whitfield, C. W., M. Band, M. F. Bonaldo, C. G. Kumar, L. Liu, et al. (2002). Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. *Genome Res.* **12** (4): 555–566.

Wuyts, J., Y. Van de Peer, T. Windelmans, and R. De Wachter (2002). The European database on small subunit ribosomal RNA. *Nucleic Acids Res.* **30**: 183–185.

Zaidi, R. H., Z. Jaal, N. J. Hawkes, J. Hemingway, and W. O. C. Symondson (1999). Can multiple-copy sequences of prey DNA be detected amongst the gut contents of invertebrate predators? *Mol. Ecol.* **8**: 2081–2087.

Zane, L., L. Bargelloni, and T. Patarnell (2002). Strategies for microsatellite isolation: a review. *Mol. Ecol.* **11**: 1–16.

Some Relevant Web Sites

The Web site of Joe Felsenstein, Department of Genetics, University of Washington, is a starting point for a wide array of data analysis programs of value for molecular systematics and ecology. You can obtain PHYLIP, PopG, DNAtree, and many others at: www@evolution.genetics.washington.edu/pub

The journal *Molecular Ecology* is online at: www.blacksci.co.uk/~cgilib/nlpage

A DNA microarray Web site, with many links, is at: www.gene-chips.com/

Another microarray site is at: cmgm.stanford.edu/pbrown/mguide/index.html

The *Drosophila* protocols page of the WWW Virtual Library, *Drosophila* section (www.ceolas.org/VL/fly/protocols.html) contains links to protocols from several laboratories.

Affymetrix's GeneChip *Drosophila* Genome is available at: www.affymetrix.com/products/drosophila_content.html

Microarray data analysis methods include GeneCluster from the Whitehead/MIT Center for Genome Research (www.genome.wi.mit.edu/MPR/software.html); Cluster and Tree View (rana.lbl.gov); and the phylogeny inference package PHYLIP (evolution.genetics.washington.edu/phylip.html)

14

Transgenic Pest and Beneficial Insects for Pest Management Programs

14.1. Overview

Genetic modification using recombinant DNA methods can now be used, almost routinely, to transform pest and beneficial arthropods. Such genetically engineered insects and mites could be used to improve future pest management programs. Goals include modifying vector mosquitoes and other insect vectors that transmit plant, human and animal diseases, so that they are unable to transmit the causal pathogens. Transgenic methods could improve genetic control programs, in which males are mass reared, sterilized by irradiation, and released to mate with wild females. Producing sterile males or producing only females by transgenic methods could improve the efficiency and effectiveness of such programs. Other goals include producing honey bees and silk moths that are disease resistant, or have other desirable economic traits. Natural enemies used in biological control programs could be modified to enhance their effectiveness, perhaps by altering their sex ratio, temperature and relative humidity tolerances, or diapause attributes.

Genetic manipulation with recombinant DNA methods requires methods for efficient and stable insertion of foreign genes into the genome of the insect and the availability of useful genes, as well as appropriate promoters and other regulatory elements to obtain effective expression of the inserted gene in both space and time. Transgenic insect strains developed for these projects should be contained in the laboratory with effective procedures until permits have been obtained from appropriate regulatory authorities that would allow their release into the environment.

Risk assessments must be conducted prior to releasing transgenic insects into the environment for either short-term experiments or permanent establishment. Potential risk issues to be resolved prior to releases include whether the inserted gene(s) (trait) is stable; whether traits can be horizontally transferred to other populations or species; whether released insects will perform as expected with regard to their geographic distribution, host or prey specificity, and other biological attributes; whether released insects will have unintended environmental effects; and, in the case of short-term releases, whether the released insects can be recovered from the field sites. Risk assessments of fitness and host specificity are relatively easy to carry out in the laboratory, but evaluations of some potential risks, such as horizontal gene transfer or unintended effects on ecosystem function, are more difficult.

A number of steps are involved in a program designed to control pest insects through transgenic methods. First, the target species must be identified as a significant pest for which conventional control tactics are ineffective because genetic manipulation is usually more expensive and time consuming than other pest management approaches. The increased costs arise, in part, because genetic manipulation with recombinant DNA techniques generates concerns about risk. The genetic engineer next will want to ask: How best can our knowledge about the pest species' physiology, ecology, or behavior be used against it? How will the transgenic strain be deployed in a pest management program?

Once a target trait has been identified, it must be genetically altered using appropriate genes and regulatory sequences to ensure the new trait is expressed at the appropriate

time and in appropriate tissues. Once the modified strain has been developed, it must be evaluated in the laboratory for fitness and stability. If the ultimate deployment method requires mass rearing of very large numbers of high-quality insects, mass rearing methods and release models must be developed. Eventually the manipulated strain will be released into greenhouses or small field plots in the field for additional evaluation of stability, fitness, and efficacy.

Permission to release a transgenic insect will have to be obtained from (perhaps several) regulatory agencies. Short-term releases initially will be made into small plots, perhaps in cages. Initial releases of transgenic insects into the environment in the United States are intended to be short-term experiments, and current regulation of such releases by the U.S. Department of Agriculture requires the researcher to retrieve all transgenic insects from the environment at the end of the experiment. At present (2002), there are no national or international guidelines regarding permanent release of transgenic insects into the environment.

Many pest management programs, especially those involving *replacement* of pest populations by the transgenic population, will require permanent establishment in the environment and may require the use of "drive" mechanisms. Several drive mechanisms, such as release of active transposable elements or *Wolbachia*, have been proposed, but an analysis of the potential risk issues associated with them has not been conducted.

14.2. Introduction

This chapter provides an overview of progress and some key questions remaining regarding the development of transgenic arthropods for use in pest management programs. The literature cited is incomplete because it is expanding rapidly, but provides the reader an entrance to this field. Box 14.1 lists some of the projects currently under way that could employ transgenic arthropods.

This chapter should make it clear that additional research and discussion regarding the potential uses of transgenic arthropods is needed. Furthermore, it is important to note that there are gaps in our regulatory framework (as of August 2002). A particular gap involves the lack of a framework for conducting risk analyses prior to the *permanent* release into the environment of transgenic insects for pest management programs.

14.3. Why Genetically Modify Insects?

14.3.1. Beneficial Insects

Domesticated and semidomesticated insects have been modified by traditional breeding methods for hundreds of years. Genetic manipulation has improved disease resistance and silk production in silk moths (Yokoyama 1979, Gopinathan 1992) and disease resistance and pollination attributes in honey bees (Rothenbuhler 1979).

Natural enemies of pest insects and mites have been modified by traditional breeding methods and by hybridization of different strains to achieve hybrid vigor (Hoy 1976, 1990a, 1993, Whitten and Hoy 1999). For example, a pesticide-resistant predatory mite, *Metaseiulus occidentalis*, developed with traditional breeding methods, was incorporated into an integrated mite management program in almonds in California (Hoy 1985).

These predators provided effective control of spider mites, reduced the need for costly pesticides, reduced production costs, and saved almond growers approximately $22 million per year, most of which was due to fewer applications of pesticides to control the pest spider mites (Headley and Hoy 1987). This project demonstrated that genetic improvement of natural enemies could result in improved pest management programs.

Genetic improvement of natural enemies for biological control of pest insects and mites by traditional genetic methods has involved selecting for resistance to pesticides, lack of diapause, and increased tolerance to temperature extremes, although modification of other traits, such as dispersal rate and sex ratio, theoretically could result in improved biological control (Hoy 1990b, 1993, Stouthamer et al. 1992).

14.3.2. Pest Insects

During the past 40 years, a number of pest insects have been sterilized by irradiation or chemicals for use in genetic control programs (Wright and Pal 1967, Pal and Whitten 1974, Curtis 1979, LaChance 1979, Whitten 1979, Tan 2000). This approach to pest management has been called the sterile insect release method (SIRM) or the sterile insect technique (SIT). Male insects are mass reared and sterilized, usually by irradiation, and released. If approximately 100 sterile males are released for each wild male, wild females should mate most often with sterile males (assuming equal fitness), resulting in reduced progeny. The SIRM program is most effective when females mate only once. Serious pests, including the Mediterranean and Caribbean fruit flies, mosquitoes, and the New World screwworm *Cochliomyia hominivorax*, have been controlled or eradicated in pest management programs (Tan 2000).

For example, the SIRM initially was used to eradicate the New World screwworm from the United States (see Chapter 10, Box 10.1). Later the program was expanded to eliminate *C. hominivorax* from Central America in order to provide a buffer zone to preclude its reintroduction into the United States. Benefits of the SIRM program in 1996 to U.S., Mexican, and Central American cattle producers were estimated to be $796 million, $292 million, and $77.9 million, respectively (Wyss 2000). The benefit to cost ratios for the eradication programs ranged from an average of 12.2 to 1 for Central America to 18 to 1 for the U.S. and Mexican programs (Wyss 2000). In addition, screwworm eradication has a significant human and wildlife health component that was not included in these calculations.

14.4. Why Use Transgenic Methods?

Traditional genetic methods have limitations, and recombinant DNA methods offer new opportunities for improving pest management programs (Table 14.1). For example, significant benefits could accrue if recombinant DNA methods allowed sterile insects to be produced without incurring the negative effects of irradiation. During the sterilization process, the insect's whole body is irradiated, which produces damage in all tissues. As a result, the SIRM requires rearing very large numbers of insects for release. Commonly, pest populations are first reduced by pesticide applications or through natural seasonal (winter) mortality so that the number of insects that have to be released can be reduced. The number of sterile males released is usually a multiple of the estimated density of wild males, with a 100 : 1 ratio of sterile to wild males typically used. Rearing huge numbers of sterile insects is costly and difficult.

Table 14.1. Pest Management Goals That Might Be Achieved with Transgenic Methods

Project type	Objective(s)	Potential method(s) (selected references)
Biological control of insect and mite pests		
Improve survival of natural enemies in environment	Enhance ability to control pests	Insert stress or pesticide resistance genes; modify diapause or temperature tolerance traits (Beckendorf and Hoy 1985, Heilmann et al. 1994, Walker et al. 1995)
Improve effectiveness of natural enemies	Alter traits that enhance effectiveness	Alter sex ratio (more females); alter longevity, fecundity, host/ prey specificity, restrict ability to fly (Beckendorf and Hoy 1985, Stouthamer et al. 1992, Heilmann et al. 1994, Walker et al. 1995)
Disease control method		
Develop insects that introduce a vaccine into their hosts when taking a blood meal	Provide low-cost vaccination against widespread, serious diseases such as malaria	Mosquito injects protein when biting their human host that elicits an immune response over time (Stowell et al. 1998)
Develop transgenic symbionts to prevent disease transmission (paratransgenesis)	Prevent transmission	Gut or other symbionts are genetically modified to eliminate a disease agent in the insect vector (Beard et al. 1993)
Improve domesticated and semidomesticated insects		
Improve silk production in silk moths (*Bombyx mori*, *Philosamia*, *Anthaerea*) in India, Japan, and China	Improve quantity, quality or type of silk	Introduce disease resistance genes into silk moths to increase production; increase number of silk genes in moths to increase yield per moth (Gopinathan 1992); introduce genes from spiders or other silk-producing arthropods into moth to produce special types of silk (Vollrath and Knight 2001)
Improve honey bees, *Apis mellifera*	Improve disease and pest resistance; improve ability to pollinate special crops; reduce aggressiveness	Insert genes for resistance to bacterial, viral, and fungal diseases and mite (*Varroa*) pests; modify pollination and aggressive behaviors
Population control		
Population replacement	Eliminate traits that make a pest by releasing genetically modified individuals that will *replace* the pest population	Eliminate ability to vector diseases (malaria, dengue, sleeping sickness, yellow fever) by altering ability of pathogen to pass gut or salivary gland barriers; eliminate need for a blood meal by vectors such as mosquitoes; alter behavior so vector feeds on only one host (if it picks up disease it can't transmit it). Replacement will require some type of *drive mechanism* or a way to select for the released population in the field (Meredith and James 1990, Higgs et al. 1998, Blair et al. 2000, James 2000, Kokoza et al. 2000)
Insert useful/deleterious genes into pest populations	Release genetically modified individuals that	Insert genes/traits into populations with a "driver" such as transposable elements,

continues

continued

Project type	Objective(s)	Potential method(s) (selected references)
	will mate with wild populations and insert genes into populations	*Wolbachia*, or "gut" symbionts, by releasing appropriate ratio of transgenic to wild insects (introgression model) (Beard et al. 1992, Kidwell and Ribeiro 1992, Richards 1993, Durvasala et al. 1997, Kiszewski and Spielman 1998, Sinkins and O'Neill 2000, Hao et al. 2001, Braig and Yan 2002)
		Release tropical mosquito population into temperate area so the hybrid progeny produced are unable to overwinter (Hanson et al. 1993).
		Release insects with active transposable elements into a population; elements cause chromosomal mutations, shut off genes or cause sterility; if the element cannot be repressed by its host, then it might cause so much damage that the population crashes (as do many laboratory populations when transposable elements invade) (Thomas et al. 2000)
Sterile insect release method (SIRM)		
Sterilize males by recombinant DNA methods	Reduce damage from irradiation or chemo-sterilization	Modify male fertility genes and regulatory sequences; activate sterility by stimuli such as light, diet, or antibiotics, allowing both sexes to be reared normally until the stimulus activates the transgene (Robinson and Franz 2000)
Mark released males with molecular marker	Discriminate between released and wild males consistently in traps	Insert benign marker that can be detected in trapped dead males consistently; green fluorescent protein may be appropriate
Develop genetic sexing method so females do not have to be reared	Reduce rearing costs Improve efficiency	Alter sex determination cascade so females can be eliminated with a conditional lethal gene during mass rearing (Heinrich and Scott 2000, Christophides et al 2001).

Recombinant DNA methods also could allow unique molecular markers, such as green fluorescent protein (GFP), to be inserted into the sterile insects, which might allow SIRM program managers to more easily discriminate between released sterile males and wild fertile males caught in the traps used to monitor the progress of the SIRM program (Handler and McCombs 2000, Higgs and Lewis 2000, Handler and Harrell 2001). Current marking methods using fluorescent dusts are not satisfactory because they can reduce fitness of the insects and the dyes do not always adhere, which could lead program managers to conclude that more wild insects are present in the field than is true.

Other significant benefits could be obtained if recombinant DNA methods make it possible to control the sex of insects being reared in SIRM programs, to introduce lethal genes or genetic loads into pest populations, or to produce vectors of human and animal diseases that are unable to transmit diseases such as malaria, dengue, yellow fever, and sleeping sickness

(Crampton et al. 1990, Meredith and James 1990, Beard et al. 1992, 1993, Collins 1994, Curtis 1994, Fryxell and Miller 1995, Evans 1996, Olson et al. 1996, Beaty and Carlson 1997, Durvasala et al. 1997, Gorman et al. 1997, Pfeifer and Grigliatti 1996, Higgs et al. 1998, Beard et al. 2000, Beaty 2000, Blair et al. 2000, Capurro et al. 2000, Collins et al. 2000, Heinrich and Scott 2000, Kokoza et al. 2000, McCarroll et al. 2000, Robinson and Franz 2000, Thomas et al. 2000, Adelman et al. 2001, Dimopoulos et al. 2001, James 2001, Yoshida et al. 2001). See Table 14.1 for an outline of the concepts involved.

Recombinant DNA techniques could make genetic improvement of beneficial insects, such as silkworms (*Bombyx mori*), honey bees (*Apis mellifera*), or natural enemies, more efficient and less expensive (Beckendorf and Hoy 1985, Walker 1989, Heilmann et al. 1994, Walker et al. 1995, Beckage 1998, Table 14.1). Once a useful gene has been cloned, it could be inserted into a number of beneficial species in a relatively short time. Furthermore, recombinant DNA methods broaden the number and type of genes potentially available for use; no longer is a project dependent upon the intrinsic genetic variability of the species under study.

Many have speculated about the role that recombinant DNA methods could play in the genetic control of insects that serve as the vectors of human and animal diseases or pests of agricultural crops. Some consider transgenic technology to be a new and vitally important pest management tool for the control of serious pests that cannot be controlled by any other means. Others have expressed reservations about the goals and methods suggested. For examples of different viewpoints, see Whitten (1985), Walker (1989), Crampton et al. (1990), Meredith and James (1990), Eggleston (1991), Fallon (1991), Handler and O'Brochta (1991), Besansky and Collins (1992), Kidwell and Ribeiro (1992), Collins (1994), Collins and Besansky (1994), Curtis (1994), Gwadz (1994), Hoy (1993, 1995, 2000a,b), Spielman (1994), Durvasala et al. (1997), Ashburner et al. (1998), Curtis and Townson (1998), Beaty (2000), Blair et al. (2000), Collins et al. (2000), Curtis (2000, 2001), James (2000, 2001), Robinson and Franz (2000), Spielman et al. (2002).

There are limitations to transgenic methods at present that require additional research. For example, traits primarily determined by single major genes are most appropriate for manipulating insects by recombinant DNA techniques at present. Methods for manipulating and stabilizing traits that are determined by complex genetic mechanisms are not yet feasible with insects, although such methods could be developed using procedures developed by plant molecular geneticists as models.

14.5. What Is Involved in a Project Using Recombinant DNA Methods?

Genetic manipulation by recombinant DNA techniques involves several steps and requires substantial investments of time and resources to answer the questions in Table 14.2. A successful project outcome probably will require that we have a thorough knowledge of the biology, ecology and behavior of the target species. Identifying one or more specific traits that, if altered, potentially would achieve the goals of the project is a critically important planning step (Table 14.2). Suitable genes must be identified and cloned, and appropriate regulatory sequences must be identified so that the inserted gene will be expressed at appropriate levels in the correct tissues and at a relevant time.

Stable transformation involves incorporating the genetic information into the germ line (ovaries and testes) so that the new genetic information is transmitted to succeeding

generations. Several approaches have been successful in inserting foreign DNA into the genome of insects (O'Brochta and Atkinson 1997, Handler 2001, Table 14.3). Several transformed lines are developed and evaluated to determine which are most fit and stable in the laboratory; such evaluations should be conducted in a manner to preclude accidental release into the environment by using appropriate containment conditions and procedures (Hoy et al. 1997, Young et al. 2000, USDA-APHIS Web site 2001). If laboratory tests indicate that the transgenic strain is fit and the trait is stably and appropriately expressed, the transgenic strain(s) may be evaluated in greenhouses or small field plots to confirm their efficacy and fitness under more natural conditions (Table 14.2, Figure 14.1).

The goals of genetic manipulation of pest and beneficial insects are different, although most of the steps are similar (Table 14.2). The goal is to *reduce* or eliminate the pest population, or its impact; by contrast, the goal is to *enhance* or increase the population of beneficial arthropods, or their impact, in the field (Hoy 2000a,b).

Table 14.2. Questions to Answer When Developing a Genetic Manipulation Project If It Is to Be Deployed Successfully

PHASE I. Defining the problem and planning the project

- What genetic trait(s) limit effectiveness of beneficial species or might reduce damage caused by the pest?
  - Do we know enough about the biology, behavior, genetics, and ecology of the target species to answer this question?
  - Is the potential trait determined by single or multiple genes?
- Can alternative control tactics be made to work more effectively and inexpensively than genetic manipulation projects, and are they more environmentally friendly?
  - The costs of genetic manipulation projects are high, and the time to develop a functional program can be quite long.
  - Transgenic technology may not be appropriate if traditional genetic or other control methods can be used because issues surrounding risk assessment of releasing transgenic arthropods into the environment for permanent establishment have not been resolved.
- How will the genetically manipulated strain be deployed?
  - Will releases be inoculative and some type of selection or drive system used to *replace* the wild strain?
  - Will the desired genes be *introgressed* (introduced) into the wild population? What selection mechanism will be used?
  - Will augmentative releases of very large numbers be required?
  - Will multiple releases be required over many years?
- What risk issues, especially of transgenic strains, should be considered in planning?
  - If pesticide resistance genes are used as a selectable marker for beneficial species, is there a possibility of the resistance gene moving to a pest?
  - What is known about the potential for horizontal gene transfer?
  - If transposable element or viral vectors are used in the transformation process, what risks might they pose if the transgenic strain is released into the environment?
  - What health or other hazards might be imposed on human subjects if the transgenic strain were released?
- What advice do the relevant regulatory authorities give regarding your plans to develop a transgenic strain?
  - Which agencies are relevant to consult for your project?

PHASE II. Developing the genetically manipulated strain and evaluating it in the laboratory

- Where will you get your gene(s)?
  - Should the transgene(s) sequence be modified to optimize expression in the target species if it is from a species with a different codon bias?

continues

Table 14.2. continued

- Is it important to obtain a high level of expression in particular tissues or life stages?
  - Where can you get the appropriate regulatory sequences?
- How can you maintain or restore genetic variability in your selection or transgenesis program?
  - Because both artificial selection and transgenic methods typically involve substantial inbreeding to obtain pure lines, how will you outcross the manipulated strain with a field population to improve its adaptation to the field or otherwise increase genetic variability?
- What methods can you use to evaluate "fitness" in artificial laboratory conditions that will best predict effectiveness in the field?
  - Have life table analyses and laboratory studies of the stability of the trait under no selection been correlated with efficacy in the field?
  - Is it possible to carry out competitive population cage studies?
- Do you have adequate containment methods to prevent premature release of the transgenic strain into the environment?
  - Have these containment methods been reviewed by appropriate regulatory authorities?
- Do you have adequate rearing methods developed for carrying out field tests?
  - Are artificial diets available to reduce rearing costs?
  - Are quality control methods available to maintain quality during mass rearing?
- What release rate will be required to obtain the goals you have set?
  - Do you have an estimate of the absolute population density of the target species in your field test?
  - What release model are you applying: inundative, inoculative, introgression, complete population replacement?
- Have you tested for mating biases, partial reproductive incompatibilities or other population genetic problems?
- If the strain is transgenic, have you obtained approval from the appropriate regulatory authorities to release the strain into the greenhouse or small plot?
  - Can you contain it in the release site?
  - Can you retrieve it from the release site at the end of the experiment?
  - Can you mitigate if unexpected problems arise?
- How will you measure effectiveness of the modified strain in the field trials?

PHASE III. Field evaluation and eventual deployment in practical pest management project

- If the small-scale field trial results obtained in Phase II were promising, questions remain to be asked prior to the deployment of the manipulated strain.
  - Are mass rearing methods adequate?
  - Is the quality control program in place?
  - Is the release model feasible?
  - Were there unexpected reproductive incompatibilities between the released and wild populations?
- If permanent releases are planned, have all the risk issues been evaluated?
- How will the program be evaluated for effectiveness?
- Will the program be implemented by the public or private sector?
- What will the program cost and what are the benefits?
- What inputs will be required to maintain the effectiveness of the program over time?

Modified from Hoy (2000a).

14.6. What Germ-Line Transformation Methods Are Available?

Inserting cloned DNA into insects can be accomplished using several different techniques (Table 14.3). If the inserted DNA is incorporated into the chromosomes in the cells that give

Table 14.3. Gene Transfer Methods for Developing Transgenic Arthropods

Method of inserting DNA	Stable or transient transformation	Host range species, order(s)
Gene targeting		
Homologous recombination	Stable	*D. melanogaster*, Diptera; not yet used with other insects
Targeted gene replacement	Stable	*D. melanogaster*, Diptera; not yet used with other insects
Modify microbial symbionts (paratransgenesis) of insects		
Gut symbionts	Stable	Symbionts chosen have narrow host range
Transposable element vectors		
Hermes	Stable	From *Musca domestica*, Diptera; transformed *D. melanogaster*, *Aedes aegypti*, *Tribolium castaneum*, *Ceratitis capitata*, *Stomoxys calcitrans*, Diptera and Coleoptera
hobo	Stable	From *D. melanogaster*, Diptera; transformed *D. melanogaster* and *D. virilis*, Diptera
mariner	Stable	Widespread; transformed *D. mauritiana*, *M. domestica*, *Bombyx mori* cells, Diptera and Lepidoptera
Minos	Stable	From *Drosophila hydei*, Diptera; Transformed *C. capitata* and *Anopheles stephensi*, Diptera
piggyBac	Stable	From *Tricoplusia ni*, Lepidoptera; Transformed several Diptera, Lepidoptera and Coleoptera
Viral vectors		
Baculoviral vectors	Transient	Used to express proteins in insects or insect cell cultures; primarily Lepidoptera, but can integrate into mammal chromosomes; efforts to produce stable transformation are underway
Pantropic retroviral vectors Engineered with broad host range	Stable (goal) Transient now	Evaluate genes and regulatory elements; very broad host range due to modification, including humans
Parvoviridae vectors	Transient Stable in future	Host range is limited to mosquitoes (?)
Polydnavirus vectors	Stable	Transforms lepidopteran and coleopteran cells in culture
Retroviruses and retrotransposons	Potentially stable	Modify *gypsy* to be a vector?
Sindbis virus vectors	Transient	Evaluate genes and regulatory elements; virus host range primarily *Culex* and *Aedes* mosquitoes and birds; horizontal transmission via water and cannibalism; vertical transmission via venereal transmission (?)

rise to the ovaries and testes, the foreign genetic material could be transmitted faithfully and indefinitely to successive generations (stable germ-line transformation).

Initial research on stable transformation methods was accomplished with *Drosophila melanogaster* when it was discovered that the *P* element could be genetically manipulated to serve as a vector to carry foreign genes into the chromosomes of germ-line cells (Rubin and Spradling 1982, Spradling and Rubin 1982; see also Chapter 9). The genes carried by the *P*-element vector became stably integrated into the chromosomes of *D. melanogaster* and were expressed.

14.6.1. *P*-Element Vectors

In molecular genetics, vectors are self-replicating DNA molecules that transfer a DNA segment between host cells. After the pioneering research of Rubin and Spradling (1982), *P*-element vectors were investigated as possible gene vectors for insects other than *Drosophila*, but generally they failed to function in insects outside the Drosophilidae (O'Brochta and Handler 1988, Handler 2000, Atkinson et al. 2001; see also Chapter 9). During this early work, mosquitoes (*Aedes aegypti*, *Anopheles gambiae*, and *Ae. triseriatus*) were transformed, but the rate of transformation was low (less than 0.1% of the microinjected embryos) and there is no evidence that the transformation was mediated by the *P* element (Morris et al. 1989, Miller et al. 1987, McGrane et al. 1988). As a result, a variety

A)



Figure 14.1. A) A transgenic strain of the predatory mite *Metaseiulus occidentalis* (Acari: Phytoseiidae) containing a *lac*Z marker gene first was released into the field on April 10, 1996, in Gainesville, Florida. This was a short-term release designed to evaluate the fitness of the strain, its stability, and our ability to predict its behavior and to contain it. Because this predatory mite lacks wings and tends to stay on the release plants if provided adequate prey, these predators were released into the center row of potted plants with the outside row of pesticide-treated plants serving as traps to reduce the likelihood of escape from the plot. The white poles surrounding the release site contain clear sticky panels (not visible) at two heights to monitor any movement of the predators out of the plot; only 2–3 were collected on the sticky panels over the course of the experiment. At the end of the experiment, the plants were placed in plastic garbage bags and autoclaved to preclude the transgenic predators from persisting in the environment. This predator, originally from the western United States, is unable to persist in Florida's hot, wet summers and so was climatically contained and could not permanently establish. Furthermore, no wild-type population was present with which it could interbreed. B) Prior to making the release, approval was obtained from the University of Florida's biosafety committee, the Florida Department of Agriculture and Consumer Services, the United States Department of Agriculture—Animal and Plant Health Inspection Service, the Florida Department of the Environment, and the U.S. Fish and Wildlife Service. Personnel from these agencies were present at the release site to ensure that the requirements of the permit were met.

B)



Figure 14.1. *continued*

of other methods for achieving transformation have been evaluated (Atkinson et al. 2001, Table 14.3).

14.6.2. Other Transposable Element Vectors

TEs such as *Hermes*, *hobo*, *mariner*, *Minos*, and *piggyBac* have been isolated from insects and genetically modified for use as vectors to transform insects other than *D. melanogaster* (Table 14.3). The process of evaluating such vectors typically involves first using transient assays to determine whether the vector functions in insect cells in cell culture or in transiently transformed insects (Handler 2001). Each vector is described briefly, including its source and potential host range, which may be relevant for risk assessment.

14.6.2.1. *Hermes*

Hermes was discovered in *Musca domestica* (O'Brochta et al. 1996) and is known to function in four dipteran families (Drosophilidae, Calliphoridae, Tephritidae and Muscidae) (Atkinson and O'Brochta 2000). Cell lines of *Anopheles gambiae* were stably transformed by *Hermes* (Zhao and Eggleston 1998), and *Hermes* transposed in embryos of *Aedes aegypti* (Sarkar et al. 1997b).

Two strains of *M. domestica* exhibited low rates of transformation (Sarkar et al. 1997a), perhaps because *Hermes* is endemic in housefly populations and some form of genetic resistance to *Hermes* has been selected for. Likewise, the sheep blowfly strain (Calliphoridae) tested exhibited low levels of transposition, perhaps because a *Hermes*-like element (*hermit*) is present in the blowfly genome. *Stomoxys calcitrans* (O'Brochta et al. 2000), *Ceratitis capitata* (Michel et al. 2001), and *Aedes aegypti* (Jasinskiene et al. 1998, 2000) have been transformed with *Hermes*. However, integrations of *Hermes* into *Ae. aegypti* did not occur

precisely, suggesting that *Hermes* could have integrated into the genome by an abnormal mechanism (Jasinskiene et al. 2000).

14.6.2.2. *hobo*

hobo, isolated from *D. melanogaster*, can function as a vector in several drosophilids (Handler and Gomez 1996), in lepidopteran (*Trichoplusia ni* and *Helicoverpa zea*) cells (DeVault et al. 1994, 1996), as well as in several tephritids (*Anastrepha suspensa*, *Bactrocera dorsalis*, *B. cucurbitae, C. capitata*, and *Toxotrypana curvicauda*) (Handler and Gomez 1996).

The stability and effectiveness of the transformed insect strains developed for pest management programs depend on the stability of the inserted gene within the genome. As a result, it is desirable to examine the genome of the population undergoing transformation to be sure that endogenous elements related to the vector are lacking before conducting transformation experiments with a TE vector. For example, tephritid flies transformed with *hobo* vectors were unstable, and excision was stimulated by heat shocks that presumably elicited the production of an endogenous *hobo*-like transposase (Atkinson et al. 1993). The excision rate was 8- to 10-fold higher than that seen for hosts lacking endogenous transposase.

It appears that it is insufficient to examine the target insect for endogenous versions of *hobo*. Laboratory assays indicated that *hobo* transposase functioned with both *hobo* and *Hermes* substrates (Sundararajan et al. 1999). On the other hand, *Hermes* transposase rarely was able to excise *hobo* elements from plasmids.

The TE family that includes *hobo* includes elements from plants, fungi, fish, insects, and humans. Thus, it appears to move horizontally relatively easily. For example, *hobo* appears to have invaded *D. melanogaster* populations after 1960 (Bonnivard et al. 2000). The broad host range makes *hobo* desirable as a vector for inserting transgenes into insects, but could be considered a negative attribute from the point of view of risk assessment when transgenic insects are evaluated prior to their release into the environment.

14.6.2.3. *mariner*

The *mariner* element initially was isolated from *Drosophila mauritiana*, but is extremely widespread among arthropods (Robertson 1993, 1995, Robertson and Lampe 1995) and is present in many other organisms, including nematodes (Grenier et al. 1999, Leroy et al. 2000), flatworms (Garcia-Fernandez et al. 1995) and hydras (Robertson 1997). This TE is found in mammals, including humans (Auge-Gouillou et al. 1995, Oosumi et al. 1995, Robertson and Martos 1997). A *mariner* vector was used to transform the chicken (Sherman et al. 1998), the zebrafish (Fadool et al. 1998) and the protozoan *Leishmania major*, (Gueiros-Filho and Beverley 1997).

The broad host range of *mariner* raises the question as to whether there is a risk that active *mariner* (or other TEs) purposefully released in insect control programs as “drivers” could invade the human or other genomes. At least two different subfamilies of *mariner* have been isolated from the human genome, which suggests our genome was invaded more than once. In fact, Oosumi et al. (1995) suggested that *mariner* vectors could be used to transform human cells, and Plasterk et al. (1999) engineered a *mariner* element to make it more active in human cells. The element (*Sleeping Beauty*) had 25-fold higher levels of activity in human cells than the “standard” *mariner*. However, the presence of ancient and degenerated *mariner* elements in the human genome could indicate that humans have developed resistance to invasion by *mariner*, suggesting the risk could be low.

To date, rates of transformation of arthropods with *mariner* vectors have been low (Lampe et al. 2000, Coates et al. 1995, Wang et al. 2000), perhaps because many insect species have, over evolutionary time, developed the ability to suppress its damaging effects. The *mariner* vector was shown to transpose in embryos of *Aedes aegypti* and to transform it (Coates et al. 1998) as well as *Musca domestica* (Yoshiyama et al. 2000a). The *mariner* element could be made more effective as an arthropod vector through genetic manipulation (Lampe et al. 1999).

14.6.2.4. *Minos*

Minos was discovered in *Drosophila hydei* and can transpose in cells and embryos of *D. melanogaster*, *Bombyx mori*, and *Anopheles stephensi* (Catteruccia et al. 2000a, Klinakis et al. 2000a, Shimizu et al. 2000) and also produce stable germ-line transformation (Loukeris et al. 1995, Catteruccia et al. 2000b). *Minos* can transform human cell lines (Klinakis et al. 2000b), which could elicit concerns about risks if *Minos* is used to transform insects destined for release into the field in pest management programs. To reduce potential risk with transgenic insects used in pest management programs, it may be necessary to eliminate *Minos* vector sequences (and other TE vector sequences) after transformation of the insect line even if the element has been "disabled." See the discussion below on "conversion," which indicates that, under some circumstances, even disabled vectors can become active.

14.6.2.5. *piggyBac*

The *piggyBac* element was isolated from a nucleopolyhedrosis virus infecting cell cultures of the moth *Trichoplusia ni* (Fraser 2000). It transposes in a cell line of another moth (Fraser et al. 1995, Elick et al. 1996) and in embryos of *D. melanogaster*, *Ae. aegypti*, and *An. gambiae* (Grossman et al. 2000) and the cabbage looper *Trichoplusia ni* (Lobo et al. 1999). In addition, *Ceratitis capitata* (Handler et al. 1998), *Pectinophora gossypiella* (Peloquin et al. 2000a), *Bombyx mori* (Tamura et al. 2000), and *Tribolium castaneum* (Berghammer et al. 1999) were transformed with a *piggyBac* vector, suggesting it has a broad host range.

14.6.3. Viral Vectors

Several types of viruses have been modified to serve as vectors in insects or insect cells (Burns 2000, Carlson et al. 2000, Olson 2000, Terzian et al. 2000, Webb 2000). In some cases, these viruses are intended to yield stable transformation, but others are expected to result in a short-term transformation of the infected tissues (Table 14.3).

14.6.3.1. Baculovirus Vectors

Nuclear polyhedrosis viruses (NPV), or baculoviruses, have double-stranded, circular DNA genomes contained within a rod-shaped protein coat. Baculoviruses infect a number of pest insects, and several have been used as biological pesticides, including *Autographa californica* NPV and *Lymantria dispar* NPV (Shuler et al. 1994). These, and the *Bombyx mori* NPV, have been exploited as vectors to carry exogenous DNA into insect cells (Miller 1988, Iatrou and Meidinger 1990, Yu et al. 1992, Yamao et al. 1999). Baculovirus vectors are used to produce a high level of commercial proteins in insect cell cultures (Frommer and Ninnemann 1995, Jones and Morikawa 1996).

The host range of baculovirus vectors has been found to include human liver cells, suggesting that baculoviruses could be used for gene therapy in humans (Hofmann et al. 1995). Recombinant baculoviruses recently were found to integrate into Chinese hamster ovary chromosomes in cell cultures, suggesting they could be used as gene vectors for transforming mammals in a stable manner (Merrihew et al. 2001).

14.6.3.2. Densonucleosis Virus Vectors

Densonucleosis viruses (Parvoviridae) are linear single-stranded DNA molecules that apparently are restricted to arthropods. Densovirus vectors are being used to deliver genes into mosquitoes for laboratory studies of gene expression. Densoviruses also might be used for biological control programs directed against mosquitoes (Beaty and Carlson 1997) and may become useful for stable transformation in the future (Carlson et al. 2000). The *Aedes aegypti* densonucleosis virus has a relatively restricted host range (*Aedes*, *Culex*, and *Culiseta* mosquitoes) (Beaty and Carlson 1997, Carlson et al. 2000).

14.6.3.3. Pantropic Retroviral Vectors

Retroviral vectors have been developed by genetically modifying the Moloney murine leukemia virus so that it contains the G envelope protein from vesicular stomatitis virus (Burns 2000). These retroviral vectors have a very wide host range (are pantropic), but are considered to be stable once inserted into the host genome because they lack the genetic information needed to propagate themselves.

These vectors are being considered for human gene therapy and can be used to transform fish, cows, clams, and amoebae, as well as lepidopteran and dipteran cells (Burns 2000). These viruses have been used to study promoter function and regulation in insect cells (Matsubara et al. 1996, Jordan et al. 1998, Burns 2000). Silkworm embryos infected with pseudotyped retroviral particles carrying the green fluorescent protein gene construct yielded larvae that contained these viral vector sequences, indicating the virus integrated into the genome (Komoto et al. 2000).

14.6.3.4. Polydnaviral Vectors

These are multisegmented DNA viruses found only in the female reproductive tracts of some hymenopteran wasps. Female wasps inject the viruses when they deposit eggs in a lepidopteran host, and the viruses disrupt the host immune system, making the host more suitable for the developing parasitoid.

Polydnavirus DNA can persist within the chromosomal DNA of a gypsy moth cell line, thus “transforming” it (McKelvey et al. 1996). Insect cell lines from other lepidopteran and coleopteran species have been transformed by this polydnavirus (Gundersen-Rindal et al. 1999). Thus, polydnaviruses potentially could be used as vectors to insert foreign DNA into some insects.

14.6.3.5. Retroviral and Retrotransposon Vectors

The *gypsy* element in *D. melanogaster* is infectious and is therefore a retrovirus; it is the first retrovirus to be identified in invertebrates (Kim et al. 1994, Bucheton 1995). *gypsy* normally is repressed (prevented from moving) by a gene in *Drosophila* called

flamenco. Apparently, *gypsy* elements invaded *D. melanogaster* a long time ago, and *D. melanogaster* survived the invasion because variants of the *flamenco* gene were able to suppress the activity of the invading *gypsy* (Pelisson et al. 1997). This is a fine example of the ability of insects to evolve resistance to transposable or other elements that cause genetic damage. Because *gypsy* is infectious, it could be modified as a vector to stably transform insects.

14.6.3.6. Sindbis Virus Vectors

Alphaviruses (Togaviridae) have a single-stranded RNA genome and have been genetically engineered as expression vectors (Beaty and Carlson 1997). These viruses can be grown in mammalian cells, and the viruses produced can infect either mosquitoes or mosquito cell cultures. Infection is sustained and Sindbis infection was used to express an antisense form of a dengue protein in *Aedes aegypti* adults, making the mosquitoes unable to transmit this human viral disease (Olson et al. 1996, Olson 2000). The Sindbis virus can be fed to mosquitoes, allowing expression of transgenes in the midgut (Olson et al. 2000).

14.6.4. Paratransgenesis

Transformation of symbionts of insects is called paratransgenesis. Beard et al. (1992, 1993, 2000) demonstrated that genetic engineering of insect gut symbionts is feasible by transforming a bacterial symbiont of the Chagas disease vector *Rhodnius prolixus*. The extracellular symbiont lives in the insect gut lumen and is transmitted from adult to progeny by contamination of egg shells or of food with infected feces. The symbionts can be genetically engineered and transmitted to host insects that are lacking symbionts after treatment with antibiotics. The symbionts of *Rhodnius* have been transformed with ampicillin and thiostrepton resistance genes and with genes coding for cecropin A and related pore-forming molecules (Richards 1993, Beard et al. 2000). The antibiotic resistance genes provide a selective advantage to the transgenic symbionts so they can survive antibiotics in the blood meal. The cecropin A and related molecules can make holes in membranes, perhaps leading to lysis of the Chagas disease pathogen.

Symbionts of tsetse flies (*Glossina* species), which are vectors of both animal and human African sleeping sickness, also have been transformed (Richards 1993, Cheng and Aksoy 1999, Aksoy et al. 2001). Proposals have been made to release tsetse flies carrying transgenic symbionts so the released flies could replace or outcompete native populations but fail to transmit the disease. Because the host range of these bacteria is narrow, horizontal movement of the transformed bacteria is unlikely.

Extracellular bacteria isolated from the gut of the walnut husk fly *Rhagoletis completa* was transformed with enhanced green fluorescent protein and zeomycin resistance genes (Peloquin et al. 2000b). This modified bacterium, if ingested, could express and deliver proteins into the gut that could enhance the nutrition of the flies, improving their vigor and competitiveness for genetic control programs using sterile insects.

14.6.5. FLP-Mediated Recombination

The ability to introduce cloned genes into the germ line at a predictable chromosomal site is especially desirable, because it reduces the likelihood of position effects on gene expression (Table 14.3). Genes introduced by TE and viral vectors insert more or less

randomly into the chromosomes, making it difficult to predict how well the transgene will be expressed.

One method for accomplishing precise insertion is based on a system found in the yeast *Saccharomyces cerevisiae*. A gene for yeast recombinase, FLP recombinase, and two inverted recombination target sites (FRTs) that are specifically recognized by the FLP recombinase have been cloned. The FLP-FRT system has been modified to insert DNA into a specific site in a *Drosophila* chromosome (Konsolaki et al. 1992, Simpson 1993, Golic et al. 1997). If the FRT sites can be inserted into other insects, the system could reduce concerns about unstable transformation that are elicited by TE and other vectors.

Because a stable FRT site must be present in the genome, a number of different lines carrying FRT sites in different chromosomal locations will have to be evaluated to determine which site permits better expression of the introduced genes. The FRT system is introduced into *D. melanogaster* using *P*-element vectors, so vector sequences may have to be removed to preclude subsequent movement.

14.6.6. No Vectors

A few experiments have delivered linear or circular plasmid DNA into the genome of insects without using a specific vector (Walker 1989, Presnail and Hoy 1992, 1996). This approach has the advantage of eliminating potential risks of introducing vector sequences into the insect genome, which could result in increased stability of the inserted genes in the genome. It assumes that the inserted gene is no more likely than any other gene to be moved by "wild" TEs or viruses.

14.7. Current and Potential Methods to Deliver Foreign DNA into Arthropod Tissues

A variety of methods have been evaluated for delivering genes and vectors into the insects in order to achieve transformation (Atkinson et al. 2001, Table 14.4). Current methods include microinjection of TE vectors and other vectors into dechorionated insect eggs or microinjection of plasmids directly into the testes of males or the abdomen of female mites or insects (maternal microinjection). Less frequently, DNA has been delivered by soaking eggs in DNA, using sperm to carry foreign DNA into eggs of the honey bee, using microprojectiles (gene gun technology) to insert DNA into insect eggs, electroporation of DNA into insect eggs, and transplanting nuclei and cells. Transformation of an insect microbial gut symbiont (called paratransgenesis because the insect genome has not been modified) has been used for tsetse fly and the vector of Chagas disease. Future methods might involve inserting artificial chromosomes into the insect genome, especially if multiple genes are to be inserted.

14.8. What Genes Are Available to Insert?

Cloned DNA can be isolated from the same or other species. It is technically feasible to insert genes from microorganisms into arthropods and have the DNA transcribed and translated, although coding sequences isolated from microorganisms must be attached to promoters (controlling elements) and other regulatory sequences derived from a eukaryotic

Table 14.4. Current and Potential Methods to Deliver Foreign DNA into Arthropod Tissues

Method of delivering DNA	Example(s) (selected references)
Artificial chromosomes	
Insert genes into artificial chromosome, insert chromosome into genome	Not yet achieved with insects but feasible with yeast and mice (Peterson et al. 1997)
Biolistic Methods	
"Gene gun"	*D. melanogaster* (Baldarelli and Lengyel 1990) *Anopheles gambiae* eggs (Mialhe and Miller 1994) *Bombyx mori* salivary glands (Horard et al. 1994)
Electroporation	
Electric current punches holes in membranes, letting DNA in	*Bombyx mori* eggs (Shamila and Mathavan 1998) *D. melanogaster*, transient expression (Kamdar et al. 1992) *Helicoverpa zea*, *Musca domestica* (Leopold et al. 1996)
Microinject eggs after dechorionation	
Method originating with *D. melanogaster* (Santamaria 1986), modified for each egg type	*Bombyx mori* (Nikolaev et al. 1993, Nagaraju et al. 1996) *Musca domestica* (Yoshiyama et al. 2000b) *Pectinophora gossypiella* (Peloquin et al. 1997)
Microinject abdomens of females	
Maternal microinjection	*Metaseiulus occidentalis* (Presnail and Hoy 1992) *Cardiochiles diaphaniae* (Presnail and Hoy 1996)
Microinject testes	*Bombyx mori* (Shamila and Mathavan 1998)
Nuclear transplantion	*D. melanogaster* (Zalokar 1981) Chimeric larvae of honey bee produced (Omholt et al. 1995)
Sperm-mediated transformation	
Insert DNA into genome via artificial Insemination	*In vitro* association of DNA with sperm (Atkinson et al. 1991) *Apis mellifera* (Robinson et al. 2000)

organism so that the gene can be expressed in insects. The regulatory sequences determine when a gene will be transcribed, at what level, in what tissues, and how long the messenger RNA can be used for translation. Considerable research is under way to identify regulatory sequences that regulate genes in specific insect tissues, such as the salivary glands and gut (Angelichio et al. 1991, Bhadra et al. 1997, Cheng and Aksoy 1999, Coates et al. 1999, Box 14.1).

It also may be possible to isolate a gene from the species being manipulated, alter it, and reinsert it into the germ line, although this approach has not yet been attempted in insects other than *Drosophila*. This approach has received increased interest in plant breeding using recombinant DNA methods because it decreases concerns about risks.

Box 14.1 lists some of the projects being conducted and genes that are being evaluated. For example, increased freeze resistance in frost-susceptible beneficial insects may be increased by gene transfer. Antifreeze protein genes cloned from the wolffish *Anarhichas lupus* and the winter flounder *Pleuronectes americanus* have been expressed in transgenic *Drosophila* (Rancourt et al. 1990, 1992, Peters et al. 1993, Duncker et al. 1995, 1996). Increased dosage of antifreeze genes resulted in improved levels of expression in *D. melanogaster* (Duncker et al. 1999). Although additional work is required to obtain beneficial insects that are able to tolerate cold temperatures, the results suggest that subtropical or tropical species of natural enemies could be modified to survive in a much broader range of climates.

Box 14.1. Some Current Research Relevant to the Control of Pest, or Improvement of Beneficial Arthropods Using Recombinant DNA Methods

Project goals	Species	Selected reference(s)[a]
Genome sequencing/analysis		
–Linkage and physical maps integrated	*Aedes aegypti*	Brown et al. 2001
–Quantitative trait loci (QTL) analysis of susceptibility to *Plasmodium gallinaceum* (bird malaria) or *Brugia*	*Ae. aegypti*	Morlais and Severson,[a] Severson 1994, Beerntsen et al. 1995
–QTL analysis of encapsulation of malaria	*Anopheles gambiae*	Gorman et al. 1997
–Expressed sequence tags from ovary, larval, and pupal cDNA libraries analyzed for gene expression	*Ae. aegypti*	Gao et al.[a]
–Bacterial artificial chromosome (BAC) library constructed	*Bombyx mori*	Wu et al. 1999
–BAC library analysis Genome project organized	*An. gambiae*	Roth et al.[a] Collins et al. 2000
–QTL analysis of refractoriness and insecticide resistance	*An. gambiae*	Zheng et al. 1997 Ranson et al.[a]
–Analysis of differentially expressed genes during infection	*Culicoides sonorensis* that transmits epizootic hemorrhagic disease of deer	Campbell et al.[a]
–Genome sequenced completely	*D. melanogaster*	Adams et al. 2000
–Genetic variability in insects able to transmit disease	Variability of various species reviewed	Gooding 1996
Marker genes for transformation		
–Enhanced green fluorescent protein (EGFP) with PAX-6 promoter	Appears to function in many insect species	Wimmer et al.[a] Pinkerton et al. 2000
	Allows transformed insects to be identified easily	
	Could serve as a marker in insects released into the environment for SIRM projects	
–Eye color genes as selectable markers	Selectable markers in insects other than *Drosophila* markers	Sarkar and Collins 2000, Zwiebel et al. 1995, White et al. 1996, Cornel et al. 1997, Ke et al. 1997
–Green fluorescent protein (GFP)	Marker genes for transgenic insects	Higgs and Lewis 2000, Handler and Harrell 2001
–Insecticide/antibiotic resistance genes	Potential markers for transgenic insects	ffrench-Constant and Benedict 2000
Models to drive genes into wild pest populations		
–Female-specific lethal system	Introduce refractoriness into vector populations	Sinkins and Hastings[a]
–Transposable elements	Invasion of populations by TEs	Brookfield 1996, Brookfield and Badge 1997, Badge and Brookfield 1997

Box 14.1. continued

Project goals	Species	Selected reference(s)[a]
–Selfish DNA elements	Drive genes into populations	Hastings 1994, Pfeifer and Grigliatti 1996, Kiszewski and Spielman 1998, Tsitrone et al. 1999
–*Notch* mutant from *D. melanogaster*	Cause dominant cold-sensitive lethality in heterozygous embryos of *D. melanogaster*	Fryxell and Miller 1995
–Release of insects carrying a dominant lethal (RIDL) (modified SIRM)	Released insects carry a conditional, dominant, sex-specific gene that can be made lethal in the laboratory but not in the field; *D. melanogaster* used as a model	Thomas et al. 2000
–Release of pesticide-resistant natural enemies	*M. occidentalis, Trioxys pallidus* Partial premating isolation, metapopulation dynamics	Hoy 1995, Caprio et al. 1991 Caprio and Hoy 1994, 1995
Tissue- or stage-specific regulation of transgenes		
–Repressors, enhancer–promoter specificity, chromatin modification	*Drosophila* (model insect)	Keller et al.[a]
–Ecdysteroid response	*Drosophila*	Brodu et al.[a] Riddiford[a]
–Stage- and tissue-specific genes elicited by a blood meal in mosquitoes	*Ae. aegypti*	Raikhel,[a] Moreira et al. 2000, Coates et al. 1999
–Male-specific promoter to develop genetic sexing tool for sterile insect releases	Promoter isolated from *D. melanogaster* functions in males of *Ceratitis capitata*	Christophides et al. 2001
–Use sex-specific promoter to drive expression of a repressible transcription factor that controls expression of a selective lethal gene product	RIDL system in *D. melanogaster* to eliminate females reared in laboratory (genetic sexing) Males sterile without irradiation	Thomas et al. 2000
Traits to modify		
–Chemosensory receptors	*Drosophila* (model)	Warr et al.[a]
–Cold tolerance	Insert fish antifreeze protein into *D. melanogaster*	Duncker et al. 1995, 1996, Walker et al. 1995
–Deliver vaccines	Use mosquitoes to inject proteins to induce immune response to diseases, such as malaria, in humans	Stowell et al. 1998, Karras et al. 1998
–Gene silencing obtained with *Hermes* and *piggyBac* TE vectors containing dsRNA	*D. melanogaster*, *Bactrocera tryoni*, *Helicoverpa armigera*	Whyard et al.[a]
–Heat shock response	*D. melanogaster*	Feder et al. 1997
–Increase life span	*D. melanogaster*	Tower 2000
–Sex determination	*D. melanogaster* (model)	Baker[a]

Box 14.1. continued

Project goals	Species	Selected reference(s)[a]
	D. melanogaster, repressible system for males only	Heinrich and Scott 2000
	An. gambiae	Pannuti et al. 2000
	Ceratitis capitata	Robertson et al.[a]
	Lucilia cuprina	Scott et al.[a]
Ability to transmit malaria and other diseases[c]		
–Anti-circumsporozoite protein prevents invasion of mosquito salivary glands	*Ae. aegypti* *Plasmodium gallinaceum* and yellow fever, as model	James et al.[a] Capurro et al. 2000
–SG1 peptide binds to salivary glands/midgut interrupting malaria transmission	*An. gambiae*	Gosh et al.[a]
–Immunotoxin in bacteria inhibit malaria parasite	Transformation of *E. coli* with antibody that kills *Plasmodium berghei* in mosquito midgut	Yoshida et al. 2001
–Salivary gland genes	Various blood-feeding mosquitoes (*An. stephensi, An. gambiae, An. arabiensis, Ae. aegypti, Culex quinquefasciatus*) and sandflies (*Phlebotomus papatasi, Lutzomyia longipalpis*)	Valenzuela et al.[a]
–Innate immunity against protozoal and viral diseases	*An. gambiae*, *Ae. aegypti, Glossina palpalis*	Luna and Zheng[a] Schneider,[a] Barrillas-Mury et al. 2000
	Immunity to malaria in *An. gambiae*	Dimopoulos et al. 1997, 2001
–Melanization immunity in mosquitoes to reduce disease transmission	*Armigeres subalbatus* (vector of filarial worm *Brugia malayi*) *An. gambiae* (vector of several malaria species) Isolation of prophenoloxidase 1 gene in *An. gambiae*	Christensen et al.,[a] Romans,[a] Shiao et al.[a] Ahmed et al. 1999, Zheng 1997
–Encapsulation of invading pathogens	*Pseudoplusia includens* (model lepidopteran) Role of polydnavirus in suppressing immunity	Strand et al.,[a] Beck and Strand[a]
–Dominant, repressible lethal genetic control system	*D. melanogaster* (model)	Thomas et al.[a]
–Resistance to dengue virus using complementary sense and antisense RNA	*Culex* or *Aedes* mosquitoes Mosquito cell lines resistant to dengue virus	Olson[a] Travanty et al.[a]
–Resistance to malaria/filaria by defensin	*An. gambiae* *Ae. aegypti* resistance to *Brugia malayi*	Eggleston et al. 2000 Vizioli et al. 2001, Lowenberger et al. 1996
–Resistance to yellow fever virus	*Ae. aegypti* resistant with antisense RNA	Higgs et al. 1998
–Regulation of oxidative stress after blood meal	*An. gambiae*	Charles et al.[a]

Box 14.1. continued

Project goals	Species	Selected reference(s)[a]
Genes expressed in salivary and gut tissues of tsetse fly	*Glossina morsitans morsitans*	Hao et al. 2001, Li et al. 2001, Yan et al. 2001
Transformation methods (stable and transient)		
–*Autographa californica* nuclear polyhedrosis virus vector	*Bombyx mori*	Mori et al. 1995
–Baculovirus-mediated dsRNA gene silencing	Lepidoptera	Chouinard et al.[a]
–Baculovirus vectors for gene expression	Expression of proteins/vaccines in cultured insect cells Adaptation to expression in mammalian cells	Kost and Condreay 1999
–Densovirus (parvovirus) vectors for cell transformation	Isolated from German cockroach and other insects; Analysis of promoters in lepidopteran densovirus; transform *Ae. aegypti* with *Aedes* densovirus	Mukha and Schal[a], Bossin and Shirk[a] Carlson et al. 2000, Afanasiev et al. 1999
–Pantropic retroviral vectors	Vector based on Moloney murine leukemia virus contains envelope glycoprotein of vesicular stomatis virus Wide host range, including Insects (review) *Anopheles*, *Aedes*, and *Drosophila* somatically transformed	Burns 2000, Jordan et al. 1998
–Polydnavirus vectors	General review; *Lymantria dispar* cell lines	Webb 2000, McKelvey et al. 1996
–Retroviral vectors (insects)	Potential vectors for transfer of genes into insects	Terzian et al. 2000
–Sindbis infection for transient assays of viral interference	Interrupt virus transmission by mosquitoes Chimeric virus with enhanced infection rate by feeding	Blair et al. 2000, Olson 2000, Seabaugh et al. 1998
–Site-specific recombination	Use of FLP-FRT system in insects other than *Drosophila*	Rong and Golic 2000
–Symbiont transformation	Transform gut symbionts to modify disease transmission	Aksoy et al. 2001, Beard et al. 2000, Cheng and Aksoy 1999, Durvasala et al. 1997
–Targeted transformation	Homologous recombination in insects	Eggleston and Zhao 2000
Transposable element vectors		
Hermes	*Ae. aegypti* transformed; *Ae. aegypti* turns on the anti-bacterial *defensin* gene after a blood meal	Jasinskiene et al. 1998, 2000 Kokoza et al. 2000
	Culex quinquefasciatus transformed with GFP	Allen et al.[a]
	Ceratitis capitata transformed	Michel et al. 2001
	Transformed 11 dipterans and 1 lepidopteran (review)	Atkinson and O'Brochta 2000

Box 14.1. continued

Project goals	Species	Selected reference(s)[a]
mariner	*Ae. aegypti* transformed with purified transposase	Coates et al. 1998, 2000
	Review of recent results	Lampe et al. 2000
	M. domestica transformed	Yoshiyama et al. 2000a
Minos	*An. stephensi* transformed with GFP	Catteruccia et al.[a]
	An. gambiae cells and embryos assayed	Catteruccia et al. 2000a
	Bombyx mori embryos assayed	Shimizu et al. 2000
piggyBac	*Ceratitis capitata*	Handler et al. 1998
	Tribolium castaneum	Berghammer et al. 1999
	D. melanogaster	Handler and Harrell 1999
	An. gambiae cells and embryos assayed	Grossman et al. 2000
	Ae. aegypti transformed with *defensin/cinnabar*	Ahmed et al.,[a] Lobo et al.[a]
	Review of transformations in other insects	Fraser 2000
	Bombyx mori transformed	Tamura et al. 2000
	Bactrocera dorsalis transformed	Handler and McCombs 2000
	Pectinophora gossypiella transformed with GFP	Peloquin et al. 2000a
	Anastrepha suspensa transformed	Handler and Harrell 2001
Wolbachia endosymbionts	Potentially useful for many arthropod species?	Kidwell and Ribeiro 1992, Sinkins and O'Neill 2000

[a]Abstracts from the Keystone Symposium, *Genetic Manipulation of Insects*, Taos, New Mexico, February 2001, organized by A. S. Raikhel, A. A. James, B. M. Christensen, R. ffrench-Constant and D. O'Brochta; these abstracts are not in the References Cited section.

Altering longevity of beneficial arthropods might result in more effective biological control of pests in some environments (Tower 2000). Research on mechanisms of aging may provide useful genes for modifying longevity of beneficial insects (Clancy et al. 2001, Tatar et al. 2001). A cloned catalase gene inserted into *D. melanogaster* provided resistance to hydrogen peroxide, which is implicated in cell damage, although the life span of transgenic flies was not prolonged (Orr and Sohal 1992). Overexpression of heat shock protein 70 has been evaluated in *D. melanogaster* as a mechanism for increasing longevity, but extra copies of heat shock protein 70 may not be beneficial under all circumstances (Feder 1999, Silbermann and Tatar 2000). A number of studies report deleterious effects of large amounts of heat shock proteins, perhaps because the production of excess proteins consumes energy and substrate so that other functions are affected negatively (Feder 1999).

Other traits that might be important or useful to introduce into beneficial insects could include shortening developmental rate, enhancing progeny production, altering sex ratio, extending temperature and relative humidity tolerances, and altering host or habitat preferences (Hoy 1976, Beckage 1998). The availability of the complete sequence of the *D. melanogaster* genome increases our ability to identify interesting genes in other arthropods using sequence similarity of conserved regions.

For pest insects, it might be desirable to insert genes that slow development, reduce their ability to overwinter or survive adverse weather, skew the sex ratio to produce a preponderance of nonvectoring males, or reduce their ability to vector pathogens (Beerntsen et al. 2000) or their propensity to take blood meals (Box 14.1).

Many of these traits probably are determined by multiple genes, which make them difficult to manipulate at present. Inserting traits determined by multiple genes into an insect by recombinant DNA methods has not been achieved, but there are several methods by which several foreign genes can be introduced into a transgenic plant that might be adapted for use in insects (Halpin et al. 2001, Hunt and Maiti 2001).

14.9. Why Are Regulatory Signals Important?

Genes consist of coding segments that determine the amino acid sequences in the enzyme or structural proteins produced. However, whether a coding region is transcribed and translated in a specific tissue is determined by a number of regulatory sequences in the DNA, including promoters and enhancers. Some of these regulatory structures are in close proximity to the coding region, while others may be located farther away. The stability of messenger RNA is influenced by signals in the RNA, which can influence the amount of protein produced. It is crucial to obtain expression of the inserted gene at appropriate times and levels, and in the targeted tissues.

A factor that may be important in maintaining the inserted DNA in the transgenic insect colony over time is the presence of origins of replication that regulate DNA replication of the chromosomes. If a transgene is inserted into a region of the chromosome far from a site where an origin of replication occurs naturally, the gene could be lost over time because it is not replicated.

Regulatory sequences from insects can be combined with a protein-coding sequence from a prokaryote such as *E. coli* to form a DNA construct that will function in an insect. However, regulatory sequences from prokaryotes do not function in insects. Because regulatory sequences may vary from species to species, the source of regulatory sequences chosen may be as important as, or even more important than, the source of the protein-coding sequences. Furthermore, some regulatory sequences allow genes to be expressed only in particular tissues or in response to particular stimuli (such as heat shock), whereas other genes are expressed in most tissues most of the time. If it is important that the inserted gene function in a tissue- or stimulus-specific manner, it is essential to identify tissue- or stimulus-specific promoters.

Currently, the number of suitable regulatory sequences available for genetic manipulation of arthropods is somewhat limited, although extensive research is being conducted to identify more, especially in mosquitoes (Morris et al. 1995, Lu et al. 1997, Li et al. 2001, Box 14.1). The heat shock (*hsp70*) promoter from *Drosophila* has been used extensively, as have the *Drosophila* actin 5C, the α1-tubulin, and the metallothionein (*Mtn*) promoters (Angelichio et al. 1991, Kovach et al. 1992, Zhao and Eggleston 1999). Modifiers are known that can cause overexpression of genes in *D. melanogaster* females (Bhadra et al. 1997).

Project goals will dictate what type of regulatory sequences might be most useful. In some cases, a low constitutive production of transgenic proteins will be useful, while in other cases high levels of protein production will be required after inducement by a specific cue. Researchers may have to evaluate the trade-offs between high levels of protein production

and the subsequent effect these have on fitness of the transgenic arthropod strain based on the specific goals of each program.

The presence of introns sometimes may be necessary for high levels of transgene expression in transgenic insects (Duncker et al. 1997). In addition, the specific location of the introduced gene can affect its level of expression, so different transformed strains must be evaluated to obtain one that expresses the trait at the desired level.

14.10. How Are Transformed Insects Identified?

After inserting the desired genes and regulatory elements, the next issue is how to detect whether the gene has in fact been incorporated into the germ line of the insect. Because transformation methods remain relatively inefficient, a screening method is needed to rapidly and consistently identify transformed individuals. This process is relatively simple in *Drosophila*, where there is a wealth of genetic information, and visible markers, such as eye-color genes, can be used to identify transgenic individuals. Most pest and beneficial arthropods lack such extensive genetic information or markers.

Identifying transformed individuals could be achieved using a pesticide resistance gene as the selectable marker (ffrench-Constant and Benedict 2000). However, the release of pesticide-resistant pest arthropods into the environment would create concerns about risk. Researchers working with the Mediterranean fruit fly have resolved to forego using resistance genes as selectable markers because of concerns about risks (Ashburner et al. 1998). Concerns about releasing resistant natural enemies might be lower, except that there is a possibility of horizontal movement of pesticide resistance genes from beneficial to pest species. No one knows at present how to quantify this potential risk.

Another option is to use antibiotic resistance genes as selectable markers to identify transgenic insects. However, horizontal movement of antibiotic resistance genes from insects into microbes in the environment could result in increased levels of antibiotic resistance in pathogenic microbes; the likelihood of this potential risk has not been measured. Antibiotic resistance gene markers are no longer considered safe for release into the environment in transgenic crops and methods have been developed to remove them (Yoder and Goldsbrough 1994, Ebinuma et al. 2001, Matthews et al. 2001). It is probably desirable to eliminate unneeded marker genes from insects prior to their permanent release into the environment.

Another potential marker is based on the β-galactosidase gene (the *lacZ* construct) isolated from *E. coli*, which can be detected by an assay that produces a blue color in the transformed insects and mites. This construct has been present in a number of organisms released into the environment, and risks associated with the release of this construct are considered low (Hoy 2000a,b). Eye color (Zwiebel et al. 1995, White et al. 1996, Cornel et al. 1997, Ke et al. 1997, Sarkar and Collins 2000) and green fluorescent protein (GFP) genes (Higgs and Lewis 2000, Horn et al. 2000, Pinkerton et al. 2000) also are considered to be safe selectable markers if transgenic insects are to be released into the environment. Unfortunately, transgenic insects with mutant eye-color genes may exhibit abnormal behavior, which could reduce their effectiveness in the field. The effects of GFP on vision could be important when the GFP gene is expressed in the eyes of insects (Horn et al. 2000). Normal behavior often is crucial to the function of released insects in pest management programs.

14.11. How to Deploy Transgenic Pest and Beneficial Arthropods

A critically important step is consideration of *how* to employ the transgenic strain in pest management programs (Table 14.2, Phase III). Ideally, the questions outlined in Table 14.5 should be considered when *initiating* the project, because genetic manipulation projects of beneficial or pest insects are not rapid, inexpensive, or simple.

The efficacy of a "drive" mechanism (such as *Wolbachia* or active transposable elements) combined with a "driven gene" to control a pest population has not yet been demonstrated in any practical pest management program. Although some small-scale experimental releases have occurred (with nontransgenic insects), this type of pest population manipulation raises a number of questions regarding risk and effectiveness. Clearly, the driver and the gene to be driven should be strongly linked if the combined system is to spread through a population (Curtis 2000, Braig and Yan 2002).

Genetic control (SIRM or SIT) programs usually require repeated releases of large numbers of sterile pest insects. The insects produced must be free of diseases, vigorous, competitive, and free of genetic deterioration caused by inbreeding or inadvertent

Table 14.5. Some Risk Issues Relevant to Releases of Transgenic Insects into the Environment

A. Attributes of the unmodified organism

- What is the origin of the transgenic organism (indigenous or nonindigenous) in the accessible environment?
- What is the insect's trophic level (predator, parasite, plant feeder) and host range?
- What other ecological relationships does it have?
- How easy is it to monitor and control it?
- How does it survive during periods of environmental stress?
- What is the potential for gene exchange with other populations?
- Is the insect involved in basic ecosystem processes?

B. Attributes of the genetic alteration

- What is the intent of the genetic alteration?
- What is the nature and function of the genetic alteration?
- How well characterized is the genetic modification?
- How stable is the genetic alteration?

C. Phenotype of modified organism compared to unmodified organism

- What is the host/prey range?
- How fit and effective is the transgenic strain?
- What is the expression level of the trait?
- Has the alteration changed the organism's susceptibility to control by natural or artificial means?
- What are the environmental limits to growth or reproduction (habitat, microhabitat)?
- How similar is the transgenic strain being tested to populations previously evaluated in field tests?

D. Attributes of the accessible environment

- Describe the accessible environment, whether there are alternative hosts or prey, wild relatives within dispersal capability of the organisms, and the relationship of the site to the potential geographic range of the transgenic strain.
- Are there endangered/threatened species present that could be affected?
- Are there agents that could move the transgenic strain present in the release environment?
- Do the test conditions provide a realistic simulation of nature?
- How effective are the monitoring and mitigation plans?

Modified from Tiedje et al. 1989; USDA 1991; and from a discussion held at a conference on "Risks of Releasing Transgenic Arthropod Natural Enemies," held November 13–16, 1993, in Gainesville, Florida.

selection for laboratory adaptations (Bush 1979). Large-scale rearing of insects is difficult and expensive. If the goal is to release pest insects, such as mosquitoes, that are able to reproduce and may bite humans and domestic animals, it also will be necessary to have the cooperation of the human inhabitants of the release area (Aultman et al. 2000). Past genetic control programs involving releases of insect vectors of disease have elicited concern by inhabitants of the affected area (Pal 1974).

Will it be possible to use induced reproductive incompatibility, perhaps caused by the symbiont *Wolbachia*, to control pest insects? This prospect has been discussed since Laven (1951) observed the impact of cytoplasmic incompatibility on *Culex pipiens* populations and suggested that it could be used as a means of controlling them (Prout 1994, Sinkins and O'Neill 2000). Yen and Barr (1974) found the cause of the incompatibility in *Culex pipiens* was due to the presence of *Wolbachia*. Experiments were conducted, but the incompatibility produced was incomplete because transmission was not fully efficient (Pal 1974). However, there is hope that genetic control by this approach could become effective; O'Neill and Karr (1990) reported that *Wolbachia* causes reduced egg hatch in *Drosophila simulans* when infected males mated with uninfected females, and Turelli and Hoffman (1991) reported that *Wolbachia* spread rapidly in field populations of *D. simulans* in California in a "natural" experiment. Turelli et al. (1992) concluded that cytoplasmic incompatibility induced by *Wolbachia* "therefore provides a mechanism for introducing cytoplasmic factors into natural populations. This may eventually be useful for introducing deleterious factors into pest insect populations." Unfortunately, this is the only example in which *Wolbachia* has been shown to "sweep" through a population in the field, so it is unknown whether such sweeps will occur with other insect species and with other strains of *Wolbachia*.

In theory, reproductive incompatibility could be transferred to a population of insects lacking *Wolbachia* by microinjection of transgenic *Wolbachia*, mass rearing of the infected individuals, and release of the insects into natural populations (Sinkins and O'Neill 2000). A number of questions remain to be resolved, including whether resistance to the cytoplasmically transmitted organisms could develop in the pest insect populations if the invasion by *Wolbachia* takes a long time, whether such incompatibility will be stable, and whether the *Wolbachia* could move horizontally to nontarget insect species. Appropriate release rates of individuals containing the *Wolbachia* are critical or the drive system can fail to function (Turelli and Hoffman 1991, Johanowicz and Hoy 2000).

Other cytoplasmic factors that might decrease fitness under specific conditions include genetically altered mitochondria or viruses that would increase susceptibility to chemicals, or cytoplasmic factors that decrease resistance to temperature extremes.

Deployment of genetically manipulated insects is complicated if some form of reproductive isolation or drive mechanism cannot be provided when the goal is to obtain population replacement or character replacement. One of the reasons genetically modified predator mites have been successfully employed in pest management programs may be because these natural enemies disperse relatively slowly (Hoy 2000a,b, Figure 14.1A). Releases of pesticide-resistant, but nontransgenic, strains of natural enemies into pesticide-treated greenhouses, orchards, or vineyards provided sufficient isolation that the genetically manipulated strains were able to establish without extensive competition from, or interbreeding with, susceptible native populations (Caprio et al. 1991). Likewise, releases of a pesticide-resistant strain of the parasitoid *Aphytis melinus* into Israeli citrus groves did not involve competition or interbreeding with susceptible populations because this species was not

present in Israel. Selection for the resistant population with pesticides could provide the drive mechanism.

Predicting whether, and how, genetically modified pest or beneficial insects will establish is difficult (Hoy 2000a,b). There are at least two models that could be employed in the establishment of a genetically modified strain in situations in which a native population exists: (a) The released strain displaces the native population and replaces it (replacement model). This model assumes that relatively little interbreeding occurs between the released and native populations. (b) Alternatively, the released strain interbreeds with the native population, and a hybrid population is produced. By appropriate strong selection, perhaps with pesticide applications, the desired trait is selected for and the resultant population contains the desired gene (introgression model).

14.12. Could Gene Silencing Reduce Program Effectiveness?

There is always the risk that a transgenic insect population could be released into the field and fail to function as expected because of a phenomenon called gene silencing. Transgenic plants and mammals often inactivate multiple copies of transgenes that overexpress proteins or are otherwise abnormal (Dorer and Henikoff 1997, Wolffe 1997, Henikoff 1998, Birchler et al. 2000, Sijen and Kooter 2000). **Gene silencing** is due to systems that have evolved as a means to prevent high levels of expression of transposable elements or viruses that can cause genetic damage to their hosts. In fungi and plants, gene silencing is associated with several mechanisms, including methylation of the DNA, as well as posttranscriptional and transcriptional processes.

Multiple mechanisms of transgene silencing occur in *D. melanogaster* (Dorer and Henikoff 1994, 1997, Pal-Bhadra et al. 1999, Jensen et al. 1999). Thus, methods may have to be developed to manage transgene silencing in insects, or this phenomenon could reduce their effectiveness after release into the field. The use of insulators or boundary elements may limit gene silencing (Bell et al. 2001), and genetic elements such as histone deacetylase RPD3, which can counteract gene silencing in both *Drosophila* and yeast, also may be useful in counteracting gene silencing (De Rubertis et al. 1996).

Gene silencing might be turned into a positive attribute if specific genes in insects could be turned off. Gene silencing has been purposefully induced in *D. melanogaster* by introducing a sequence that codes for an extended hairpin-loop RNA by *P*-mediated transformation (Kennerdell and Carthew 2000). Perhaps endogenous gene expression and developmental processes could be modified in other insects by gene silencing.

14.13. Potential Risks Associated with Releasing Transgenic Arthropods

Risk equals the potential for damage *and* the likelihood of its occurrence. Risk estimates may be different for pest versus beneficial insects and may depend on whether the insect is expected to persist in the environment or is unable to reproduce and cannot persist (Table 14.5). Risks also will vary with the specific transgene(s) inserted. It is easier to suggest potential types of damage than to quantify the likelihood of its occurrence at this time.

14.13.1. Relative Risks

The least risky transgenic insect could be the domesticated silk moth (*B. mori*), which is unable to survive on its own in the wild. Transgenic *B. mori* are unlikely to have a negative effect on the environment because they should not be able to persist if they were accidentally released. Other risks might be associated with transgenic *B. mori*, however.

Transgenic pest or beneficial insects that are sterile and unable to reproduce should pose a lower risk than insects that are able to reproduce and persist in the environment. Transgenic pest or beneficial insects that are unable to persist because the environment is unsuitable during a portion of the year also are likely to pose a low risk (McDermott and Hoy 1997, Hoy 2000a,b, Figure 14.1A).

Honey bees, *Apis mellifera*, are only semidomesticated and thus can escape human management to survive in the wild. Transgenic honey bees could pose a greater environmental risk than the domesticated silk moth for this reason.

14.13.2. General Risk Issues

Evaluating the risks associated with releasing insects and mites that have been manipulated with recombinant DNA techniques will likely include, as a minimum, the questions or principles outlined in Tables 14.2 and 14.5, but other issues may become important as we learn more about risk assessment procedures (Foster et al. 2000, Kapuscinski 2002). Current concerns can be summarized as:

- Is the transgenic population stable?
- Has its host or prey range been altered?
- Does it have the potential to persist in the environment?
- Will the transgenic strain have unintended effects on other species or environmental processes?

The first three questions are relatively easy to answer with a variety of laboratory experiments (Li and Hoy 1996, McDermott and Hoy 1997, Presnail et al. 1997). The last issue is much more difficult to answer.

Releases of transgenic arthropods in the United States are now evaluated by several regulatory agencies on a case-by-case basis (Figure 14.1B). Permits are issued at present only for short-term releases in controlled situations so that unexpected outcomes might be mitigated more readily (Young et al. 2000, USDA-APHIS 2001).

Because initial releases are intended to be temporary, one question we may need to answer is, how far and how quickly can the transgenic strain disperse from the experimental release site? Can adequate containment be maintained in the short-term release site? Less is known about the dispersal behavior of many insects than might be needed. For example, Raymond et al. (1991) suggest that there has been a worldwide migration of *Culex pipiens* mosquitoes carrying amplified organophosphorus resistance genes, perhaps aided by accidental human transport. If migration is the basis for these widespread genes, then dispersal of some transgenic strains could be far more rapid and extensive than anticipated. Containment of transgenic predatory mites, which lack wings, is relatively easy (Figure 14.1A).

14.13.3. Horizontal Gene Transfer

One risk issue that is unusually difficult to quantify is the risk of horizontal transfer of transgenes, transposable elements, or *Wolbachia* to other organisms (Droge et al. 1998).

Our knowledge of horizontal transfer and TEs only began in the 1950s when Barbara McClintock discovered TEs in maize. Horizontal gene transfer could occur from one insect population to another of the same species, from one insect species to another, or to other organisms in the environment. It is difficult to quantify this risk because we lack fundamental information on the frequencies and mechanisms of horizontal gene transfer. Because horizontal movement is rare, effective sampling and statistical methods are especially important. The whole topic of horizontal gene transfer in insects has received limited scientific attention until relatively recently.

Horizontal transfer of *P* and *mariner* elements across species of *Drosophila* provides some of the best data for horizontal movement in insects. Horizontal transfer of genes does occur between insect species by movement of naturally occurring TEs (Houck et al. 1991, Plasterk 1993). Horizontal transfer is thought to be rare, yet we have observed more than one such transfer within historical times in *D. melanogaster* and may have missed other examples because we were not looking. The *P* element invaded *D. melanogaster* populations within the past 50 years, perhaps from a species in the *D. willistoni* group. *P* elements might have been transferred between these *Drosophila* species by a semiparasitic mite (Houck et al. 1991). Another TE, *hobo*, also appears to have invaded natural populations of *D. melanogaster* around the 1960s (Bonnivard et al. 2000), the second invasion of this well-studied insect in the past 40 to 50 years.

Transfer of TE vectors from transgenic insect populations to other organisms, including humans, is potentially feasible, although these transfers should occur very rarely. Recall that risk is determined by frequency of occurrence and the damage that might occur. In this case, the frequency is expected to be very low if the natural invasions represent a realistic estimate of frequency. If active TEs are purposefully released as drive mechanisms or if conversion of inactive TE vectors into active ones can occur, then the frequency could be higher.

It is difficult to estimate the potential damage invasions of TEs could have on nontarget species. For example, *mariner* is widespread and the data suggest that: 1) *mariner* elements have been present in insects for a long time, although some lineages have lost them, and 2) horizontal transfer has occurred between different insect families and orders, although some transfers occurred so long ago that many of the *mariners* are degraded and inactive, probably because of a successful defense against the damage they cause to the insect's genome (Lampe et al. 2000). Lampe et al. (2000) noted the "most recent events occur[red] at least 100,000 years ago." The two *mariner*s in the human genome probably invaded in the "past 100 million years" (Lampe et al. 2000).

We are still discovering new aspects of the evolutionary role of TEs (Kidwell and Lisch 2001), which makes it difficult to predict what would happen if insects were released that contained active TEs or inactive TE vectors (Petrov et al. 1995). The safest course might be to remove any introduced TE vector sequences from a transgenic insect strain prior to its *permanent* release into the environment to reduce the probability that the transgene will move, either within the strain or horizontally between different populations or species.

Elements other than *P* and *mariner* also move horizontally. Jordan et al. (1999) showed that a long terminal repeat retrotransposon (a different class of element than the *P* and *mariner* elements) in the *D. melanogaster* group species moved into *D. willistoni*, perhaps within the past 100 to 200 years.

Horizontal transfer of DNA could be mediated by insect viruses. The *piggyBac* element was discovered embedded within the genome of a baculovirus (Fraser 2000), and another Tc1-like transposon was found in the *Cydia pomonella* granulovirus (Jehle et al. 1998).

If horizontal transmission of transgenes by viruses were to occur in the field, there is no guarantee that genes inserted into an insect species would remain within that species.

The concern was raised that horizontal gene transfer might even occur when DNA is eaten. Although most consumed DNA is degraded, that is not always true (Schubbert et al. 1997, 1998). For example, bacteriophage DNA fed to mice can persist in fragmented form in the gut and can penetrate the intestinal wall and reach the nuclei of leukocytes, spleen, and liver cells (Schubbert et al. 1998). Fetal and newborn progeny of female mice fed such DNA during pregnancy had the phage DNA in various organs. Furthermore, the foreign DNA was located in the nuclei and associated with the chromosomes, although the DNA had not integrated (Schubbert et al. 1998). Such an association of DNA with the chromosomes could affect normal gene function. More recently, Hohlweg and Doerfler (2001) fed mice soybean leaves containing a specific gene and analyzed its fate. This experiment provided a more natural delivery system than feeding naked DNA. The results indicated that DNA in soybean leaves could be found less frequently in mice tissues than when naked DNA was fed, in part because the amount of DNA was reduced by about five orders of magnitude (Hohlweg and Doerfler 2001).

Genetic engineering of insect gut symbionts might allow the movement of the inserted genes between the many types of microorganisms found within the insect gut (Watanabe and Sato 1998, Watanabe et al. 1998). *Enterobacter cloacae*, a bacterium found in the guts of insects, and *Erwinia herbicola*, a bacterium that grows on the surface of plants, grow in the guts of silk moth larvae and exchange genetic information via plasmids at very high rates (Watanabe and Sato 1998, Watanabe et al. 1998). The bacteria containing the new genetic information were found in the feces of the insects, suggesting that this method of horizontal gene transfer is a frequent event in nature. If gut symbionts of pest insects are transformed with antibiotic resistance genes, these genes might move horizontally to other bacteria within the insect gut. Transfer of antibiotic resistance genes to pathogens could be harmful; horizontal transfer of antibiotic resistance genes has led to a serious medical crisis because some human pathogens are now resistant to almost all available antibiotics (Witte 1998).

Whether horizontal gene transfer will cause harm would certainly depend on the gene(s) transferred and its destination. The most serious harm might occur if the TE or viral vector inserted into germ-line tissues so it could be transmitted to succeeding generations. However, damage also might occur if the elements damaged somatic tissues; for example, the movement of *mariner* in the soma reduced the lifespan of *Drosophila simulans* males (Nikitin and Woodruff 1995). The movement of retroelements into human breast, colon, and testicular tissues can induce cancer or Duchenne muscular dystrophy (Capy et al. 1996).

The role of TEs in the evolution of genomes is undergoing reevaluation, and it is clear that naturally occurring horizontal gene transfer between species has provided some of the variability upon which evolution has acted (Plasterk 1993, Krishnapillai 1996, Britten 1997). It is unlikely that the presence of a transgene in an insect will increase the small probability that the transgene will be transferred to another species by horizontal transfer *if* the TE or viral vector sequences used to insert the transgene could be removed prior to release into the field. Even then, however, the probability of horizontal transfer will not be zero.

Disabled TE vectors probably pose a low risk of horizontal gene transfer of the transgene to other organisms. However, it is possible for an inactivated vector to become active through a process called conversion. Peronnet et al. (2000) showed that conversion can transform an inactive *P* element into an active one through the interaction of three different *P* elements in the genome in a three-step process. Conversion could make a transgene unstable within the transgenic insect's genome and could, in theory, pose a risk for horizontal gene transfer.

The potential risks of using TEs as drivers for inserting useful genes into insect populations should be evaluated carefully and on a case-by-case basis. As noted by Kidwell and Evgen'ev (1999),

> the transposability of mobile elements, their potential for rapid, and sometimes massive, amplification in copy number, their ability to change genomic locations, as well as their propensity for horizontal transfer, makes the generalization of results from model organisms far less reliable. Extrapolation of results from one species to another must therefore be made with caution.

14.14. Permanent Releases of Transgenic Arthropods into the Environment

Currently (2002), there are no guidelines for evaluating the risks of releasing transgenic arthropods for *permanent* establishment in the environment. Experience suggests that the probability that a new organism will become established in a new environment is small. For example, even when we are optimizing conditions in order to obtain establishment of natural enemies in classical biological control programs, only about 24% of the introduced species actually establish. Historical examples of biological invasions of pests or of establishment of classical biological control agents demonstrate a lack of predictability in such releases (Ehler 1990).

Transgenic arthropods could pose somewhat increased risks over those posed by invasive species because they are likely to be released in very large numbers and into appropriate environmental conditions; it is likely that most invasive species enter the new environment in low numbers and may not invade optimal environments. Even so, Williamson (1992) speculated that the greater the genetic novelty, the greater the possibility of surprising results of invasive species. Because transgenic insects are novel, they should be considered likely to provide surprises.

Analyses of potential risks associated with transgenic insects should include evaluation of the survival, reproduction, and dispersal of transgenic populations and their effects on other species in the community (Table 14.5). Questions also should be asked about the inserted DNA, its stability, and its possible effect on other species should the genetic material be transferred (Tables 14.2, 14.5). Both state and federal regulatory agencies in the United States, including state departments of agriculture and the U.S. Department of Agriculture, Animal and Plant Health Inspection Service (USDA-APHIS), have to be consulted for permission to release transgenic agricultural pests and natural enemies of agricultural pests. If the release occurs at a university, permission to release is required from the campus biosafety committee. Questions about potential detrimental effects of the transgenic arthropods on threatened and endangered species will be asked by state and federal agencies, including the U.S. Department of Interior, Fish and Wildlife Service (Young et al. 2000). Releases of transgenic insect vectors of human disease also raise questions regarding human consent.

14.14.1. Models to Predict

Can we use models to predict the outcome of releases of transgenic insects in pest management programs? Many types of population and genetic models could be used in attempts to predict what will happen when genetically modified insects are released into the environment in pest management programs. We do not know, however, which model types are

most likely to be predictive of the actual outcome of field releases because few models have been evaluated with empirical data (Hoy 2000a).

Predicting field results from mathematical models can be difficult. Three models were developed to predict the success of a biological control program involving applications of fungi for control of grasshoppers and locusts (Wood and Thomas 1999). All three models fit the empirical data; one predicted sustained control at low levels after a single pathogen application, but the other two predicted that repeated pathogen applications would be necessary. Analysis of these divergent expectations demonstrated that two assumptions made by ecologists and modelers are suspect: First, quantitatively similar models need not give even qualitatively similar predictions (contrary to expectations). Second, the sensitivity analysis of model predictions to parameter variation is not always sufficient to ensure the accuracy of the predictions (Wood and Thomas 1999).

Some current population models may lack key ingredients, such as partial reproductive isolation. For example, Caprio and Hoy (1995) developed a stochastic simulation model that varied the degree of mating bias between pesticide-resistant and -susceptible strains of natural enemies, diploidy state (diplo-or haplo-diploid), degree of dominance of the resistance allele, and degree to which mating biases extended to the hybrid progeny. The results were counterintuitive but demonstrated that models offer insights into the complexities of population genetics and dynamics that might be overlooked. A common assumption made in many models is that all genotypes of a species mate at random, but this assumption may mask important interactions such as mating bias or partial reproductive incompatibilities. The efficacy of transgenic insect releases could be jeopardized if mating biases exist between released and wild populations.

Empirical data generally are lacking to compare the relative usefulness of different model types in predicting insect population dynamics. Theoretical ecologists usually assume homogeneous and continuous populations. Metapopulation models, by contrast, assume that populations exist in patches varying in area, degree of isolation, and quality. Metapopulation biology increasingly is being recognized as relevant to our understanding of population ecology, genetics, and evolution (Hanski 1998). Recent data, and a variety of metapopulation models, indicate that spatial structure affects populations as much as birth and death rates, competition, and predation (Caprio and Hoy 1994).

14.15. Conclusions

Genetic manipulation projects of pest and beneficial insects share many problems and issues. Because the potential risks of permanent releases of transgenic insects into the environment have received very little evaluation, it is appropriate to release relatively low-risk examples first. This might involve the release of a transgenic beneficial insect that is carrying either a marker gene or sterile pest insects that cannot reproduce in the environment (Hoy 2000a). These initial releases of transgenic strains should not contain active TEs or viral vectors (Hoy 1992a,b; Walker et al. 1995, Ashburner et al. 1998).

Releases in the United States will involve a two-step process. Initial releases will be experimental and on a small scale. It is unknown what issues must be resolved prior to permanent releases of transgenic insects into the environment. Risk analyses will add a significant cost in both time and resources to pest management projects involving transgenic arthropods. It took about ten years for the first transgenic crop to become commercially available. It is difficult to predict how long it will be before transgenic insects are released

permanently into the environment. Risk analysis protocols have not yet been developed by other governments or international agencies.

The issue is not *if* transgenic insects should be released, but when and how? The debate over evaluation methods and risk issues should include a variety of viewpoints. Much of the debate will be parallel to the debate on the risks of introducing natural enemies for classical biological control programs into new environments (Howarth 1991, Ruesink et al. 1995). Most introduced insect natural enemies have provided significant benefits, with only a few examples of potential or demonstrated harm to the environment. Despite this, caution is warranted. Ewell et al. (1999) reported the conclusions of a workshop on the risks of deliberate introductions of species into new environments. The participants did not discriminate between the potential risks of genetically modified organisms and those of unmodified organisms introduced into new environments. Ewell et al. (1999) noted that assessing risks is complex and concluded:

> Benefits and costs of introductions [of new organisms] are unevenly distributed among ecosystems, within and across regions, among sectors of society, and across generations. Although an introduction may meet a desired objective in one area, at one time, or for some sectors of society, unwanted and unplanned effects may also occur. Introduced organisms can, therefore, simultaneously have both beneficial and costly effects. Furthermore, the relative magnitudes of costs and benefits vary both in space and over time.

Ewell et al. (1999) recommended developing a single framework for evaluating all types of introductions, noted there is a need for retrospective analyses of past introductions (of nontransgenic insects), and noted the importance of having a holistic view of the invasion process. Ewell et al. (1999) concluded:

> At the extremes, these views [of risks] range from a handful of advocates of no introductions, or of such rigorous pre-introduction proof of benignness that all introductions are effectively prohibited, to an equally small group that advocates a freewheeling global eco-mix of species. ... Most proponents of purposeful introductions understand the risks (but believe that technology can deal with them), and most conservation biologists recognize the potential benefits to be derived from carefully controlled introductions. Clearly, there is a need to bring all parties together on common ground that can lead to objective, science-based decisions by policymakers.

Science-based discussions of the potential risks associated with releases of transgenic insects and mites for pest management programs are urgently needed. Such discussion should be carried out by people with diverse viewpoints and vision. Inappropriate releases and unintended consequences could detrimentally affect all projects involving transgenic arthropods. More funding and effort should be devoted to research on risk assessment methods for transgenic arthropods (Hoy 2000b).

The potential value of transgenic arthropods to practical pest management problems is often discussed in terms of the social, public health, and economic costs associated with malaria and other arthropod-borne diseases (Crampton 1994, Curtis and Townson 1998, Blair et al. 2000, Marshall 2000). Traditional pest management programs for insect vectors of diseases have serious limitations or have failed, especially for malaria. For example, despite enormous efforts, malaria is an increasingly important health problem, with at least 400 million people falling ill with malaria each year. Between 1 and 3 million people die each year, especially children younger than 5 years of age in Africa (Marshall 2000).

Will deployment of transgenic mosquitoes unable to vector malaria contribute to a solution to malaria? Or will vaccines and drugs be developed that will mitigate the problem?

There can be no clear answers without carrying out the appropriate experiments, and different scientists have very divergent opinions as to how scarce research funds should be spent.

Miller (1989) reviewed malaria control strategies and pointed out that it is unlikely there will be "a magic bullet that will eliminate malaria." He noted that "Even DDT could not be called such a weapon, at least in retrospect" because of the problem of resistance to the pesticide. The strategy of relying on single tactics in pest management, whether it be control of mosquitoes or of agricultural pests, ultimately always has failed. Relying on transgenic insects as a sole tactic is unlikely to be an exception. For example, the complexity of genetic structure in *Anopheles gambiae* populations in west Africa (Lanzaro et al. 1998), which may be reflective of the complex genetic architecture of many other insects, suggests that release programs involving a single transgenic strain are unlikely to be successful. The integration of several compatible tactics has been more sustainable than relying upon a single pest management tactic; multitactic management of medically important disease vectors also is more likely to be sustainable (Miller 1989).

Past experience with natural enemies genetically manipulated by traditional selection methods suggests that the most readily implemented pest management projects employing transgenic natural enemies in biological control will be those where releases can be conducted in relatively small areas such as temporary cropping systems, or where the natural enemy has a low dispersal rate and can be established in individual orchards, or where the natural enemy is released into a geographic region where the wild strain does not occur. The most difficult projects to implement are likely to be those in which the new transgenic strain is expected to replace the endemic population. Projects that require the transgenic strain to replace a wild strain may require very strong selection or drivers. Teams of experts may have to develop the mass rearing technology, quality control methods, and necessary information on population structure and hidden partial reproductive isolation mechanisms that are likely to occur (Hoy 2000b).

One short-term field release of a transgenic predatory mite containing a molecular marker occurred in 1996 (Hoy 2000a). Permission to release a transgenic strain of the pink bollworm containing the gene for green fluorescent protein was requested during 2001 (USDA-APHIS 2001). It is logical to assume that releases of sterile transgenic insects (such as sterile Mediterranean fruit flies that contain a marker transgene) could occur with relatively low risk because the sterile flies should be unable to establish permanently in the environment. Likewise, because *B. mori* is unable to survive on its own in the environment, the use of transgenic strains of this insect should present a relatively low risk.

There currently are no guidelines for evaluating the risks of permanent releases of transgenic insects into the environment. Nor is it clear which regulatory agency(ies) in the United States will regulate releases of transgenic insects that vector human diseases. The "Coordinated Framework" that regulates transgenic organisms has gaps, making regulation of transgenic insect vectors of human diseases problematic. Uniform regulations regarding appropriate facilities and procedures for containing transgenic arthropods prior to their release into the environment also are lacking (Hoy et al. 1997).

Other questions remain. For example, is informed consent of humans in the release zone required before transgenic mosquitoes are released that are unable to transmit human diseases? Who bears the burden of liability should environmental, or other, harm occur after releasing transgenic insects? Most transgenic insects are being developed by scientists working for nonprofit organizations such as universities and governmental agencies, but commercial development of transgenic insects has begun. For example, a company has been

formed to develop mosquitoes that transmit a protein in their saliva when they bite humans that elicits an immune response; others have been formed to develop transgenic insects that produce proteins that could be used as drugs or vaccines.

Significant and rapid advances have been achieved in transformation of insects and the identification of potentially useful genes to insert, although knowledge about useful genes for improving the effectiveness of natural enemies for biological control programs remains limited. However, new opportunities should arise over the next few years to identify useful genes now that the *Drosophila* Genome Project has been completed (Adams et al. 2000).

There is considerable controversy as to the role that transgenic mosquitoes could play in controlling or eradicating diseases such as malaria. One view is presented by Curtis (2000):

> There is much excitement about transgenesis as a way to generate strains of mosquito that cannot transmit malaria. If a single dominant gene with these properties could be engineered, this would be an improvement on *Plasmodium* nonsusceptible strains that have already been selected by old-fashioned breeding techniques. However, without extremely reliable systems for driving the transgenes into wild vector populations, possession of a nontransmitter strain would be of no practical use. Even if a totally reliable gene-driving system were produced, there might well be strong political objections to the irrevocable release of genetically manipulated insects that bite people.

Deploying a transgenic insect in a pest management program is an awesome challenge, requiring risk assessments, detailed knowledge of the population genetics, biology, and behavior of the target species under field conditions, as well as coordinated efforts between molecular and population geneticists, ecologists, regulatory agencies, pest management specialists, and sustained efforts to educate the public about the benefits and potential risks of releasing transgenic insects into the environment. Risks associated with releasing arthropod vectors of human disease require evaluation by vector-borne disease specialists, ethicists, and funding and regulatory agency officials (Aultman et al. 2000). When the mosquito or other vector of human or animal disease is transgenic, such risk assessments may be especially controversial.

References Cited

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, and P. G. Amanatides, et al. (2000). The genome sequence of *Drosophila melanogaster. Science* **287**: 2185–2195.

Adelman, Z. N., C. D. Blair, J. O. Carlson, B. J. Beaty, and K. E. Olson (2001). Sindbis virus-induced silencing of dengue viruses in mosquitoes. *Insect Mol. Biol.* **10**: 265–273.

Afanasiev, B. N., T. W. Ward, B. J. Beaty, and J. O. Carlson (1999). Transduction of *Aedes aegypti* mosquitoes with vectors derived from *Aedes* densovirus. *Virology* **257**: 62–72.

Ahmed, A., D. Martin, A. G. O. Manetti, S. J. Han, W. J. Lee, K. D. Mathiopoulos, H. M. Muller, F. C. Kafatos, A. Raikhel, and P. T. Brey (1999). Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **96**: 14797–14800.

Aksoy, S., I. Maudlin, C. Dale, A. S. Robinson, and S. L. O'Neill (2001). Prospects for control of African trypanosomiasis by tsetse vector manipulation. *Trends Parasitol.* **17**: 29–35.

Angelichio, M. L., J. A. Beck, H. Johansen, and M. Ivey-Hoyle (1991). Comparison of several promoters and polyadenylation signals for use in heterologous gene expression in cultured *Drosophila* cells. *Nucleic Acids Res.* **19**: 5037–5043.

Ashburner, M., M. A. Hoy, and J. Peloquin (1998). Transformation of arthropods-research needs and long term prospects. *Insect Mol. Biol.* **7**(3): 201–213.

Atkinson, P. W., and D. A. O'Brochta (2000). *Hermes* and other *hAT* elements as gene vectors in insects, pp. 219–235. In: *Insect Transgenesis. Methods and Applications.* A. M. Handler and A. A. James, Eds. CRC Press, Boca Raton, FL.

Atkinson, P. W., E. R. Hines, S. Beaton, K. I. Matthaei, K. C. Reed, and M. P. Bradley (1991). Association of exogenous DNA with cattle and insect spermatozoa in vitro. *Mol. Reprod. Dev.* **23**: 1–5.

Atkinson, P. W., W. D. Warren, and D. B. O'Brochta (1993). The *hobo* transposable element of *Drosophila* can be cross-mobilized in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90**: 9693–9697.

Atkinson, P. W., A. C. Pinkerton, and D. A. O'Brochta (2001). Genetic transformation systems in insects. *Annu. Rev. Entomol.* **46**: 317–346.

Auge-Gouillou, C., Y. Bigot, N. Pollet, M. H. Hamelin, M. Meunier-Rotival, and G. Periquet (1995). Human, and other mammalian genomes contain transposons of the *mariner* family. *FEBS Lett.* **368**: 541–546.

Aultman, K. S., E. D. Walker, F. Gifford, D. W. Severson, C. B. Beard, and T. W. Scott (2000). Managing risks of arthropod vector research. *Science* **288**: 2321–2322.

Badge, R. M., and J. F. Y. Brookfield (1997). The role of host factors in the population dynamics of selfish transposable elements. *J. Theor. Biol.* **187**: 261–171.

Baldarelli, R. M., and J. A. Lengyel (1990). Transient expression of DNA after ballistic introduction into *Drosophila* embryos. *Nucleic Acids Res.* **18**: 5903–5904.

Barillas-Mury, C., B. Wizel, and Y. S. Han (2000). Mosquito immune responses and malaria transmission: lessons from insect model systems and implications for vertebrate innate immunity and vaccine development. *Insect Biochem. Mol. Biol.* **30**: 429–442.

Beard, C. B., P. W. Mason, S. Aksoy, R. B. Tesh, and F. F. Richards (1992). Transformation of an insect symbiont and expression of a foreign gene in the Chagas' disease vector *Rhodnius prolixus*. *Amer. J. Trop. Med. Hyg.* **46**: 195–200.

Beard, C. B., S. L. O'Neill, R. B. Tesh, F. F. Richards, and F. Aksoy (1993). Modification of arthropod vector competence via symbiotic bacteria. *Parasitol. Today* **9**: 179–183.

Beard, C. B., R. V. Durvasula, and F. B. Richards (2000). Bacterial symbiont transformation in Chagas disease vectors, pp. 289–303. In: *Insect Transgenesis. Methods, and Applications*. A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Beaty, B. J. (2000). Genetic manipulation of vectors: a potential novel approach for control of vector-borne diseases. *Proc. Natl. Acad. Sci. USA* **97**: 10295–10297.

Beaty, B. J., and J. O. Carlson (1997). Molecular manipulation of mosquitoes. *Curr. Opin. Infect. Dis.* **10**: 372–376.

Beckage, N. E. (1998). Parasitoids, and polydnaviruses. *Bio Science* **48**: 305–311.

Beckendorf, S. K., and M. A. Hoy (1985). Genetic improvement of arthropod natural enemies through selection, hybridization or genetic engineering techniques, pp. 167–187. In: *Biological Control in Agricultural IPM Systems.* M. A. Hoy, and D. C. Herzog, Eds. Academic Press, New York.

Beerntsen, B. T., D. W. Severson, J. A. Klinkhammer, V. A. Kassner, and B. M. Christensen (1995). *Aedes aegypti*: A quantitative trait locus (QTL) influencing filarial worm intensity is linked to QTL for susceptibility to other mosquito-borne pathogens. *Exp. Parasitol.* **81**: 355–362.

Beerntsen, B. T., A. A. James, and B. M. Christensen (2000). Genetics of mosquito vector competence. *Microbiol. Mol. Biol. Rev.* **64**: 115–137.

Bell, A. C., A. G. West, and G. Felsenfeld (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**: 447–450.

Berghammer, A. J., M. Klingler, and E. A. Wimmer (1999). A universal marker for transgenic insects. *Nature* **402**: 370–371.

Besansky, N. J., and F. H. Collins (1992). The mosquito genome: organization, evolution and manipulation. *Parasitol. Today* **8**: 186–192.

Bhadra, U., M. Pal-Bhadra, and J. A. Birchler (1997). A trans-acting modifier causing extensive overexpression of genes in *Drosophila melanogaster*. *Mol. Gen. Genet.* **254**: 621–634.

Birchler, J. A., M. Pal Bhadra, and U. Bhadra (2000). Making noise about silence: repression of repeated genes in animals. *Curr. Opin. Genet. Dev.* **10**: 211–216.

Blair, C. D., Z. N. Adelman, and K. E. Olson (2000). Molecular strategies for interrupting arthropod-borne virus transmission by mosquitoes. *Clin. Microbiol. Rev.* **13**: 651–661.

Bonnivard, E., C. Bazin, B. Denis, and D. Higuet (2000). A scenario for the *hobo* transposable element invasion, deduced from the structure of natural populations of *Drosophila melanogaster* using tandem repeats. *Genet. Res. Camb.* **75**: 13–23.

Braig, H. R., and G. Yan (2002). The spread of genetic constructs in natural insect populations, pp. 252–314. In: *Genetically Engineered Organisms. Assessing Environmental and Human Health Effects*. D. K. Letourneau, and B. E. Burrows, Eds. CRC Press, Boca Raton, FL.

Britten, R. J. (1997). Mobile elements inserted in the distant past have taken on important functions. *Gene* **205**: 177–182.

Brookfield, J. F. Y. (1996). Models of the spread of non-autonomous selfish transposable elements when transposition and fitness are coupled. *Genet. Res. Camb.* **67**: 199–209.

Brookfield, J. F. Y., and R. M. Badge (1997). Population genetics models of transposable elements. *Genetica* **100**: 281–294.

Brown, S. E., D. W. Severson, L. A. Smith, and D. L. Knudson (2001). Integration of the *Aedes aegypti* mosquito genetic linkage and physical maps. *Genetics* **157**: 1299–1305.

Bucheton, A. (1995). The relationship between the *flamenco* gene and *gypsy* in *Drosophila*: how to tame a retrovirus. *Trends Genet.* **11**: 349–353.

Burns, J. C. (2000). Pantropic retroviral vectors for insect gene transfer, pp. 125–137. In: *Insect Transgenesis. Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Bush, G. L. (1979). Ecological genetics and quality control, pp. 145–152. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Caprio, M. A., and M. A. Hoy (1994). Metapopulation dynamics affect resistance development in a predatory mite. *J. Econ. Entomol.* **87**: 525–534.

Caprio, M., and M. A. Hoy (1995). Premating isolation in a simulation model generates frequency-dependent selection and alters establishment rates of resistant natural enemies. *J. Econ. Entomol.* **88**: 205–212.

Caprio, M. A., M. A. Hoy, and B. E. Tabashnik (1991). A model for implementing a genetically improved strain of the parasitoid *Trioxys pallidus* Haliday (Hymenoptera: Aphidiidae). *Am. Entomol.* **34**: 232–239.

Capurro, M. D. L. J. Coleman, B. T. Beerntsen, K. M. Myles, K. E. Olson, E. Rocha, A. U. Krettli, and A. A. James (2000). Virus-expressed, recombinant single-chain antibody blocks sporozoite infection of salivary glands in *Plasmodium gallinaceum*-infected *Aedes aegypti. Am. J. Trop. Med. Hyg.* **62**: 427–433.

Capy, P., C. Bazin, D. Anxolabehere, and T. Langin (1996). Horizontal transfers and the evolution of transposable elements, pp. 15–30. In: *Transgenic Organisms and Biosafety*. E. R. Schmidt, and T. Hankeln, Eds. Springer-Verlag, Berlin.

Carlson, J., B. Afanasiev, and E. Suchman (2000). Densonucleosis viruses as transducing vectors for insects, pp. 138–159. In: *Insect Transgenesis. Methods, and Applications*. A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Catteruccia, F., T. Nolan, C. Blass, H. M. Muller, A. Crisanti, F. C. Kafatos, and T. G. Loukeris (2000a). Toward *Anopheles* transformation: *Minos* element activity in anopheline cells and embryos. *Proc Natl. Acad. Sci. USA* **97**: 2157–2162.

Catteruccia, F., T. Nolan, T. G. Loukeris, C. Blass, C. Savakis, F. C. Kafatos, and A. Crisanti (2000b). Stable germline transformation of the malaria mosquito *Anopheles stephensi. Nature* **405**: 959–962.

Cheng, Q., and S. Aksoy (1999). Tissue tropism, transmission and expression of foreign genes *in vivo* in midgut symbionts of tsetse flies. *Insect Mol. Biol.* **8**: 125–132.

Christophides, G. K., C. Savakis, A. C. Mintzas, and K. Komitopoulou (2001). Expression, and function of the *Drosophila melanogaster* ADH in male *Ceratitis capitata* adults: a potential strategy for medfly genetic sexing based on gene-transfer technology. *Insect Mol. Biol.* **10**: 249–254.

Clancy, D. J., D. Gems, L. G. Harshman, S. Oldham, H. Stocker, E. Hafen, S. J. Leevers, and L. Partridge (2001). Extension of life-span by loss of CHICO, a *Drosophila* insulin receptor substrate protein. *Science* **292**: 104–106.

Coates, C. J., C. L. Turney, M. Frommer, D. A. O'Brochta, and P. W. Atkinson (1995). The transposable element *mariner* can excise in nondrosophilid insects. *Mol. Gen. Genet.* **249**: 246–252.

Coates, C. J., N. Jasinskiene, L. Miyashiro, and A. A. James (1998). *Mariner* transposition and transformation of the yellow fever mosquito, *Aedes aegypti. Proc. Natl. Acad. Sci. USA* **95**: 3742–3751.

Coates, C. J., N. Jasinskiene, G. B. Pott, and A. A. James (1999). Promoter-directed expression of recombinant fire-fly luciferase in the salivary glands of *Hermes*-transformed *Aedes aegypti. Gene* **226**: 317–325.

Coates, C. J., N. Jasinskiene, D. Morgan, L. R. O. Tosi, S. M. Beverley, and A. A. James (2000). Purified *mariner (MosI)* transposase catalyzes the integration of marked elements into the germ-line of the yellow fever mosquito, *Aedes aegypti. Insect Biochem. Mol. Biol.* **30**: 1003–1008.

Collins, F. H. (1994). Prospects for malaria control through the genetic manipulation of its vectors. *Parasitol. Today* **10**: 370–371.

Collins, F. H., and N. J. Besansky (1994). Vector biology and the control of malaria in Africa. *Science* **264**: 1874–1875.

Collins, F. H., L. Kamau, H. A. Ranson, and J. M. Vulule (2000). Molecular entomology and prospects for malaria control. *Bull. World Health Org.* **78**: 1412–1423.

Cornel, A. J., M. Q. Benedict, C. S. Rafferty, A. J. Howells, and F. H. Collins (1997). Transient expression of the *Drosophila melanogaster cinnabar* gene rescues eye color in the white eye (WE) strain of *Aedes aegypti. Insect Biochem. Mol. Biol.* **27**: 993–997.

Crampton, J., A. Morris, G. Lycett, A. Warren, and P. Eggleston (1990). Transgenic mosquitoes: a future vector control strategy? *Parasitol. Today* **6**: 31–36.

Crampton, J. M. (1994). Molecular studies of insect vectors of malaria. *Adv. Parasitol.* **34**: 1–31.

Curtis, C. F. (1979). Translocations, hybrid sterility, and the introduction into pest populations of genes favorable to man, pp. 19–30. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Curtis, C. F. (1994). The case for malaria control by genetic manipulation of its vectors. *Parasitol. Today* **10**: 371–374.

Curtis, C. F. (2000). The case for deemphasizing genomics for malaria control. *Science* **290**: 1508.

Curtis, C. (2001). Present, and future control of malaria. *Science* **291**: 436.

Curtis, C. F., and H. Townson (1998). Malaria: existing methods of vector control and molecular entomology. *Brit. Med. Bull.* **54**: 311–325.

De Rubertis, F., D. Kadosh, S. Henchoz, D. Pauli, G. Reuter, K. Struhl, and P. Spierer (1996). The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**: 589–591.

DeVault, J. D., and S. K. Narang (1994). Transposable elements in Lepidoptera: *hobo*-like transposons in *Heliothis virescens* and *Helicoverpa zea. Biochem. Biophys. Res. Commun.* **203**: 169–175.

DeVault, J. D., K. J. Hughes, R. A. Leopold, O. A. Johnson, and S. K. Narang (1996). Gene transfer into corn earworm (*Helicoverpa zea*) embryos. *Genome Res.* **6**: 571–579.

Dimopoulos, G., A. Richman, H. M. Muller, and F. C. Kafatos (1997). Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc. Natl. Acad. Sci. USA* **94**: 11508–11513.

Dimopoulos, G., H. M. Muller, E. A. Levashina, and F. C. Kafatos (2001). Innate defense against malaria infection in the mosquito. *Curr. Opin. Immunol.* **13**: 79–88.

Dorer, D. R., and S. Henikoff (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila. Cell* **77**: 993–1002.

Dorer, D. R., and S. Henikoff (1997). Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans. Genetics* **147**: 1181–1190.

Droge, M., A. Puhler, and W. Selbitschka (1998). Horizontal gene transfer as a biosafety issue: a natural phenomenon of public concern. *J. Biotechnol*. **64Z**: 79–90.

Duncker, B. P., C. P. Chen, P. L. Davies, and V. K. Walker (1995). Antifreeze protein does not confer cold tolerance to transgenic *Drosophila melanogaster. Cryobiology.* **32**: 521–527.

Duncker, B. P., J. A. Hermans, P. L. Davies, and V. K. Walker (1996). Expression of a cystine-rich fish antifreeze in transgenic *Drosophila melanogaster. Transgenic Res.* **5**: 49–55.

Duncker, B. P., P. L. Davies, and V. K. Walker (1997). Introns boost transgene expression in *Drosophila melanogaster. Mol. Gen. Genet.* **254**: 291–296.

Duncker, B. P., P. L. Davies, and V. K. Walker (1999). Increased gene dosage augments antifreeze protein levels in transgenic *Drosophila melanogaster. Transgenic Res.* **8**: 45–50.

Durvasala, R., A. Gumbs, A. Panackal, O. Kruglov, S. Aksoy, R. Merrifield, F. Richards, and C. Beard (1997). Prevention of insect borne disease: an approach using transgenic symbiotic bacteria. *Proc. Natl. Acad. Sci. USA* **94**: 3274–3278.

Ebinuma, H., K. Sugita, E. Matsunaga, S. Endo, K. Yamada, and A. Komamine (2001). Systems for the removal of a selection marker and their combination with a positive marker. *Plant Cell. Rep.* **20**: 383–392.

Eggleston, P. (1991). The control of insect-borne disease through recombinant DNA technology. *Heredity* **66**: 161–172.

Eggleston, P., and Y. Zhao (2000). Targeted transformation of the insect genome, pp. 29–52. In: *Insect Transgenesis, Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Eggleston, P., W. Lu, and Y. Zhao (2000). Genomic organization and immune regulation of the defensin gene from the mosquito, *Anopheles gambiae. Insect Mol. Biol.* **9**: 481–490.

Ehler, L. E. (1990). Environmental impact of introduced biological-control agents: implications for agricultural biotechnology, pp. 85–96. In: *Risk Assessment in Agricultural Biotechnology*. Proc. Intern. Conf., J. J. Marois and G. Bruyening, eds., Univ. Calif., Div. Agric. Natur. Res. Publ. No. (1928).

Elick, T. A., C. A. Bauser, and M. J. Fraser (1996). Excision of the *piggyBac* transposable element *in vitro* is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* **98**: 33–41.

Evans, E. (1996). Insect vector control, pp. 323–338. In: *Biotechnology and Integrated Pest Management*. G. J. Persley, Ed. CAB Internat., Wallingford, UK.

Ewell, J. J., D. J. O'Dowd, J. Bergelson, C. C. Daehler, C. M. D'Antonio, L. D. Gomez, D. R. Gordon, R. J. Hobbs, A. Holt, K. R. Hopper, C. E. Hughes, M. LaHart, R. R. B. Leakey, W. G. Lee, L. L. Loope, D. H. Lorence,

S. M. Louda, A. E. Lugo, P. B. McEvoy, D. M. Richardson, and P. M. Vitousek (1999). Deliberate introductions of species: research needs. Benefits can be reaped, but risks are high. *Bio Science* **49**: 619–630.

Fadool, J. M., D. L. Hartl, and J. E. Dowling (1998). Transposition of the *mariner* element from *Drosophila mauritiana* in zebrafish. *Proc. Natl. Acad. Sci. USA* **95**: 5182–5186.

Fallon, A. (1991). DNA-mediated gene transfer: applications to mosquitoes. *Nature* **352**: 828–829.

Feder, M. E. (1999). Organismal, ecological, and evolutionary aspects of heat-shock proteins and the stress response: established conclusions and unresolved issues. *Am. Zool.* **39**: 857–864.

Feder, M. E., N. V. Cartano, L. Milos, R. A. Krebs, and S. L. Lindquist (1997). Effect of engineering *Hsp70* copy number on Hsp70 expression and tolerance of ecologically relevant heat shock in larvae and pupae of *Drosophila melanogaster. J. Exp. Biol.* **119**: 1637–1844.

ffrench-Constant, R. H., and M. Q. Benedict (2000), pp. 109–120. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Foster, K. R., P. Vecchia, and M. H. Repacholi (2000). Science and the precautionary principle. *Science* **288**: 979–981.

Fraser, M. J. (2000). The TTAA-specific family of transposable elements: identification, functional characterization, and utility for transformation of insects, pp. 249–268. In: *Insect Transgenesis. Methods, and Applications*. A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Fraser, M. J., L. Cary, K. Boonvisudhi, and H. H. Wang (1995). Assay for movement of lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **221**: 397–407.

Frommer, W. B., and O. Ninnemann (1995). Heterologous expression of genes in bacterial, fungal, animal, and plant cells. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**: 419–444.

Fryxell, K. J., and T. A. Miller (1995). Autocidal biological control: a general strategy for insect control based on genetic transformation with a highly conserved gene. *J. Econ. Entomol.* **88**: 1221–1232.

Garcia-Fernandez, J., J. R. Bayascas-Ramirez, G. Marfany, A. M. Munoz-Marmol, A. Casali, J. Baguna, and E. Slo (1995). High copy number of highly similar *mariner*-like transposons in planarian (Platyhelminthes): Evidence for a trans-phyla horizontal transfer. *Mol. Biol. Evol.* **12**: 421–431.

Gopinathan, K. P. (1992). Biotechnology in sericulture. *Curr. Sci.* **62**: 283–287.

Golic, M. M., Y. S. Rong, R. B. Petersen, S. L. Lindquist, and K. G. Golic (1997). FLP-mediated DNA mobilization to specific target sites in *Drosophila* chromosomes. *Nucleic Acids Res.* **25**: 3665–3671.

Gooding, R. H. (1996). Genetic variation in arthropod vectors of disease-causing organisms: obstacles and opportunities. *Clin. Microbiol. Rev.* **9**: 301–320.

Gorman, M. J., D. W. Severson, A. J. Cornel, F. H. Collins, and S. M. Paskewitz (1997). Mapping a quantitative trait locus involved in melanotic encapsulation of foreign bodies in the malaria vector, *Anopheles gambiae. Genetics* **146**: 965–971.

Grenier, E., M. Abadon, F. Brunet, P. Capy, and P. Abad (1999). A *mariner*-like transposable element in the insect parasite nematode *Heterorhabditis bacteriophora. J. Mol. Evol.* **48**: 328–336.

Grossman, G. L., C. S. Rafferty, M. J. Fraser, and M. Q. Benedict (2000). The *piggyBac* element is capable of precise excision and transposition in cells and embryos of the mosquito, *Anopheles gambiae*. *Insect Biochem. Mol. Biol.* **20**: 909–914.

Gueiros-Filho, F. J., and S. M. Beverley (1997). Trans-kingdom transposition of the *Drosophila* element *mariner* within the protozoan *Leishmania*. *Science* **276**: 1716–1719.

Gundersen-Rindal, D., D. E. Lynn, and E. M. Dougherty (1999). Transformation of lepidopteran and coleopteran insect cell lines by *Glyptapanteles indiensis* polydnavirus DNA. In Vitro *Cell. Dev. Biol.-Animal* **35**: 111–114.

Gwadz, R. W. (1994). Genetic approaches to malaria control: how long the road? *Am. J. Trop. Med. Hyg*. **50**(6) Suppl.: 116–125.

Halpin, C., A. Barakate, B. M. Askari, J. C. Abbott, and M. D. Ryan (2001). Enabling technologies for manipulating multiple genes on complex pathways. *Plant Mol. Biol.* **47**: 295–310.

Handler, A. M. (2000). An introduction to the history and methodology of insect gene transfer, pp. 3–26. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Handler, A. M. (2001). A current perspective on insect gene transformation. *Insect Biochem. Mol. Biol.* **31**: 111–128.

Handler, A. M., and S. P. Gomez (1996). The *hobo* transposable element excises and has related elements in tephritid species. *Genetics* **143**: 1339–1347.

Handler, A. M., and R. A. Harrell (1999). Germline transformation of *Drosophilla melanogaster* with the *piggyBac* transposon vector. *Insect Mol Biol.* **8**: 449–458.

Handler, A. M., and R. A. Harrell (2001). Transformation of the Caribbean fruit fly, *Anastrepha suspensa*, with a *piggyBac* vector marked with polyubiquitin-regulated GFP. *Insect Biochem. Mol. Biol.* **31**: 199–205.

Handler, A. M., and S. D. McCombs (2000). The *piggyBac* transposon mediates germline transformation in the Oriental fruit fly and closely related elements exist in its genome. *Insect Mol. Biol.* **9**: 605–612.

Handler, A. M., and D. A. O'Brochta (1991). Prospects for gene transformation in insects. *Annu. Rev. Entomol.* **36**: 159–183.

Handler, A. M., S. D. McCombs, M. J. Fraser, and S. H. Saul (1998). The lepidopteran transposon vector, *piggyBac*, mediates germ-line transformation in the Mediterranean fruit fly. *Proc. Natl. Acad. Sci. USA* **95**: 7520–7523.

Hanski, I. (1998). Metapopulation dynamics. *Nature* **396**: 41–49.

Hanson, S. M., J. P. Mutebi, G. B. Craig, Jr., and R. J. Novak (1993). Reducing the overwintering ability of *Aedes albopictus* by male release. *J. Am. Mosq. Contr. Assoc.* **9**: 78–83.

Hao, Z., I. Kasumba, M. J. Lehane, W. C. Gibson, J. Kwon, and S. Aksoy (2001). Tsetse immune responses and trypanosome transmission: implications for the development of tsetse-based strategies to reduce trypanosomiasis. *Proc. Natl. Acad. Sci. USA* **98**: 12648–12653.

Hastings, I. M. (1994). Selfish DNA as a method of pest control. *Proc. Trans. R. Soc. Lond. B* **344**: 313–324.

Headley, J. C., and M. A. Hoy (1987). Benefit/cost analysis of an integrated mite management program for almonds. *J. Econ. Entomol.* **80**: 555–559.

Heilmann, L. J., J. D. DeVault, R. L. Leopold, and S. K. Narang (1994). Improvement of natural enemies for biological control: a genetic engineering approach, pp. 167–189. In: *Applications of Genetics to Arthropods of Biological Control Significance*. S. K. Narang, A. C. Bartlett, and R. M. Faust, Eds. CRC Press, Boca Raton, FL.

Heinrich, J. C., and M. J. Scott (2000). A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program. *Proc. Natl. Acad. Sci. USA* **97**: 8229–8232.

Henikoff, S. (1998). Conspiracy of silence among repeated transgenes. *BioEssays* **20**: 532–535.

Higgs, S., and D. L. Lewis (2000). Green fluorescent protein (GFP) as a marker for transgenic insects, pp. 93–108. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Higgs, S., J. O. Rayner, K. E. Olson, B. S. Davis, B. J. Beaty, and C. D. Blair (1998). Engineered resistance in *Aedes aegypti* to a West African and a South American strain of yellow fever virus. *Am. J. Trop. Med. Hyg.* **58**: 663–670.

Hofmann, C., V. Sandig, G. Jennings, M. Rudolph, P. Schlag, and M. Strauss (1995). Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc. Natl. Acad. Sci. USA* **92**: 10099–10103.

Hohlweg, U., and W. Doerfler (2001). On the fate of plant or other foreign genes upon the uptake in food or after intramuscular injection in mice. *Mol. Genet. Genomics* **265**: 225–233.

Horard, B., A. Mange, B. Pelissier, and P. Couble (1994). *Bombyx* gene promoter analysis in transplanted silk gland transformed by particle delivery system. *Insect Mol. Biol.* **3**:261–265.

Horn, C., B. Jaunich, and E. A. Wimmer (2000). Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev. Genes Evol.* **210**: 623–629.

Houck, M. A., J. B. Clark, K. R. Peterson, and M. G. Kidwell (1991). Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **253**: 1125–1129.

Howarth, F. G. (1991). Environmental impacts of classical biological control. *Annu. Rev. Entomol.* **36**: 485–509.

Hoy, M. A. (1976). Genetic improvement of insects: fact or fantasy. *Environ. Entomol.* **5**: 833–839.

Hoy, M. A. (1985). Almonds: integrated mite management for California almond orchards, pp. 299–310. In: *Spider Mites, Their Biology, Natural Enemies, and Control*. Vol. 1B. W. Helle, and M. W. Sabelis, Eds. Elsevier, Amsterdam.

Hoy, M. A. (1990a). Genetic improvement of arthropod natural enemies: becoming a conventional tactic, pp. 405–417. In: *New Directions in Biological Control*. R. Baker, and P. Dunn, Eds. UCLA Symp. Mol. Cell. Biol., New Series, Vol. 112. Alan R. Liss, New York.

Hoy, M. A. (1990b). Pesticide resistance in arthropod natural enemies: variability and selection responses, pp. 203–236. In: *Pesticide Resistance in Arthropods*. R. T. Roush, and B. E. Tabashnik, Eds. Chapman, and Hall, New York.

Hoy, M. A. (1992a). Commentary: Biological control of arthropods: genetic engineering and environmental risks. *Biol. Control* **2**: 166–170.

Hoy, M. A. (1992b). Criteria for release of genetically improved phytoseiids: an examination of the risks associated with release of biological control agents. *Exp. Appl. Acarol.* **14**: 393–416.

Hoy, M. A. (1993). Transgenic beneficial arthropods for pest management programs: An assessment of their practicality and risks, pp. 357–369. In: *Pest Management: Biologically Based Technologies*. R. D. Lumsden, and J. L. Vaughn, Eds. Am. Chem. Soc. Conf. Proc. Series., Washington, DC.

Hoy, M. A. (1995). Impact of risk analyses on pest-management programs employing transgenic arthropods. *Parasitol. Today* **11**: 229–232.

Hoy. M. A. (2000a). Deploying transgenic arthropods in pest management programs: Risks and realities, pp. 335–367. In: *Insect Transgenesis. Methods, and Applications*. A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Hoy, M. A. (2000b). Transgenic arthropods for pest management programs: risks and realities. *Exp. Appl. Acarol.* **24**: 463–495.

Hoy, M. A., R. D. Gaskalla, J. L. Capinera, and C. N. Keierleber (1997). Laboratory containment of transgenic arthropods. *Amer. Entomol*. *Winter*: 206–209, 256.

Hunt, A. G., and I. B. Maiti (2001). Strategies for expressing multiple foreign genes in plants as polycistronic constructs. *In Vitro Cell. Dev. Biol. Plant* **37**: 313–320.

Iatrou, K., and R. G. Meidinger (1990). Tissue-specific expression of silkmoth chorion genes *in vivo* using *Bombyx mori* nuclear polyhedrosis virus as a transducing vector. *Proc. Natl. Acad. Sci. USA* **87**: 3650–3654.

James, A. A. (2000). Control of disease transmission through genetic modification of mosquitoes, pp. 319–333. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

James, A. A. (2001). Present, and future control of malaria. *Science* **291**: 435.

Jasinskiene, N., C. J. Coates, M. Q. Benedict, A. J. Cornel, C. S. Rafferty, A. A. James, and F. H. Collins (1998). Stable, transposon mediated transformation of the yellow fever mosquito, *Aedes aegypti*, using the *Hermes* element from the housefly. *Proc. Natl. Acad. Sci. USA* **95**: 3743–3747.

Jasinskiene, N., C. J. Coats, and A. A. James (2000). Structure of *Hermes* integrations in the germline of the yellow fever mosquito, *Aedes aegypti. Insect Mol. Biol.* **9**: 11–18.

Jehle, J. A., A. Nickel, J. M. Vlak, and H. Backhaus (1998). Horizontal escape of the novel Tc1-like Lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J. Mol. Evol.* **46**: 215–224.

Jensen, S., M. P. Gassama, and T. Heidmann (1999). Taming of transposable elements by homology-dependent gene silencing. *Nat. Genet.* **21**: 209–213.

Johanowicz, D. L., and M. A. Hoy (2000). *Wolbachia* infection dynamics in experimental laboratory populations of *Metaseiulus occidentalis. Entomol. Exp. Appl.* **93**: 259–268.

Jones, I., and Y. Morikawa (1996). Baculovirus vectors for expression in insect cells. *Curr. Opin. Biotechnol.* **7**: 512–516.

Jordan, T. V., H. Shike, V. Boulo, V. Cedeno, Q. Fang, B. S. Davis, M. Jacobs-Lorena, S. Higgs, K. J. Fryxell, and J. C. Burns (1998). Pantropic retroviral vectors mediate somatic cell transformation and expression of foreign genes in dipteran insects. *Insect Mol. Biol.* **7**: 215–222.

Jordan, I. K., L. V. Matyunina, and J. F. McDonald (1999). Evidence for recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA* **96**: 12621–12625.

Kamdar, P., G. von Allmen, and V. Finnerty (1992). Transient expression of DNA in *Drosophila* via electroporation. *Nucleic Acids Res.* **20**: 3526.

Kapuscinski, A. R. (2002). Controversies in designing useful ecological assessments of genetically engineered organisms, pp. 385–415. In: *Genetically Engineered Organisms. Assessing Environmental and Human Health Effects*. D. K. Letourneau, and B. E. Burrows, Eds. CRC Press, Boca Raton, FL.

Karras, M., I. Siden-Kiamos, J. M. Crampton, and R. E. Sinden (1998). Studies on the expression of a parasite antigen in *Drosophila melanogaster* as a model to evaluate the use of transgenic haematophagous insects as vaccine vehicles. Abstr. 211, p. 39. In: *Keystone Symposium: Toward the Genetic Manipulation of Insects*, Taos, NM.

Ke, Z., M. Q. Benedict, A. J. Cornel, N. J. Besansky, and F. H. Collins (1997). The *Anopheles albimanus white* gene: molecular characterization of the gene and a spontaneous *white* gene mutation. *Genetica* **101**: 87–96.

Kennerdell, J. R., and R. W. Carthew (2000). Heritable gene silencing in *Drosophila* using double-stranded RNA. *Nat. Biotechnol.* **18**: 896–898.

Kidwell, M. G., and M. B. Evgen'ev (1999). How valuable are model organisms for transposable element studies? *Genetica* **107**: 103–111.

Kidwell, M. G., and D. R. Lisch (2001). Transposable elements, parasitic DNA, and genome evolution. *Evol.* **55**: 1–24.

Kidwell, M. G., and J. M. C. Ribeiro (1992). Can transposable elements be used to drive disease refractoriness genes into vector populations. *Parasitol. Today* **8**: 325–329.

Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prud'homme, and A. Bucheton (1994). Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of*Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**: 1285–1289.

Kiszewski, A. E., and A. Spielman (1998). Spatially explicit model of transposon-based genetic drive mechanisms for displacing fluctuating populations of anopheline vector mosquitoes. *J. Med. Entomol.* **35**: 584–590.

Klinakis, A. G., T. G. Loukeris, A. Pavlopoulos, and C. Savakis (2000a). Mobility assays confirm the broad host-range activity of the *Minos* transposable element and validate new transformation tools. *Insect Mol. Biol.* **9**: 269–275.

Klinakis, A. G., L. Zagoraiou, D. K. Vassilatis, and C. Savakis (2000b). Genome-wide insertional mutagenesis in human cells by the *Drosophila* mobile element *Minos. EMBO Rep.* **1**: 416–421.

Kokoza, V., A. Ahmed, W. L. Cho, N. Jasinskiene, A. A. James, and A. Raikhel (2000). Engineering blood meal-activated systemic immunity in the yellow fever mosquito, *Aedes aegypti. Proc. Natl. Acad. Sci. USA* **97**: 9144–9149.

Komoto, N., C. Thibert, V. Boulo, Q. Fang, J. C. Burns, and P. Couble (2000). Gene introduction into silkworm embryos with a pseudotyped retroviral vector. *J. Seric. Sci. Jpn.* **69**: 55–61.

Konsolaki, M., M. Sanicola, T. Kozlova, V. Liu, B. Arca, C. Savakis, W. M. Gelbart, and F. C. Kafatos (1992). FLP-mediated intermolecular recombination in the cytoplasm of *Drosophila embryos. New Biol.* **4**: 551–557.

Kost, T. A., and J. P. Condreay (1999). Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Curr. Opin. Biotechnol.* **10**: 428–433.

Kovach, M. J., J. O. Carlson, and B. J. Beaty (1992). A *Drosophila* metallothionein promoter is inducible in mosquito cells. *Insect Mol. Biol.* **1**: 37–43.

Krishnapillai, V. (1996). Horizontal gene transfer. *J. Genet.* **75**: 219–232.

LaChance, L. E. (1979). Genetic strategies affecting the success and economy of the sterile insect release method, pp. 8–18. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Lampe, D. J., B. J. Akerley, E. J. Rubin, J. J. Mekalanos, and H. M. Robertson (1999). Hyperactive transposase mutants of the *Himar1 mariner* transposon. *Proc. Natl. Acad. Sci. USA* **96**: 11428–11433.

Lampe, D. J., K. K. O. Walden, J. M. Sherwood, and H. M. Robertson (2000). Genetic engineering of insects with *mariner* transposons, pp. 237–248. In: *Insect Transgenesis. Methods and Applications*. A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Lanzaro, G. C., Y. T. Toure, J. Carnahan, L. Zheng, G. Dolo, S. Traore, V. Petrarca, K. D. Vernick, and C. E. Taylor (1998). Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proc. Natl. Acad. Sci. USA* **95**: 14260–14265.

Laven, H. (1951). Crossing experiments with *Culex* strains. *Evolution* **5**: 370–375.

Leopold, R. A., K. J. Hughes, and J. D. DeVault (1996). Using electroporation and a slot cuvette to deliver plasmid DNA to insect embryos. *Gen. Anal. Biomol. Eng.* **12**: 197–200.

Leroy, H., F. Leroy, C. Auge-Gouillou, P. Castagnone-Sereno, F. Vanlerberghe-Masutti, Y. Bigot, and P. Abad (2000). Identification of *mariner*-like elements from the root-knot nematode *Meloidogyne* spp. *Mol. Biochem. Parasitol.* **107**: 181–190.

Li, J., and M. A. Hoy (1996). Adaptability, and efficacy of transgenic and wild-type *Metaseiulus occidentalis* (Acari: Phytoseiidae) compared as part of a risk assessment. *Exp. Appl. Acarol.* **20**: 563–574.

Li, S., J. Kwon, and S. Aksoy (2001). Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans. Insect Mol. Biol.* **10**: 69–76.

Lobo, N., X. Li, and M. J. Fraser, Jr. (1999). Transposition of the *piggyBac* element in embryos of *Drosophila melanogaster, Aedes aegypti* and *Trichoplusia ni. Mol. Gen. Genet.* **261**: 803–810.

Loukeris, T. G., I. Livadaras, B. Arca, S. Zabalou, and C. Savakis (1995). Gene transfer into the Medfly, *Ceratitis capitata*, with a *Drosophila hydei* transposable element. *Science* **270**: 2002–2005.

Lowenberger, C. A. M. T. Ferdig, P. Bulet, S. Khalili, J. A. Hoffmann, and B. M. Christensen (1996). *Aedes aegypti*: induced antibacterial proteins reduce the establishment and development of *Brugia malayi. Exp. Parasitol.* **83**: 191–201.

Lu, M., P. J. Farrell, R. Johnson, and K. Iatrou (1997). A baculovirus (*Bombyx mori* nuclear polyhedrosis virus) repeat element functions as a power constitutive enhancer in transfected insect cells. *J. Biol. Chem.* **272**: 30724–30728.

Marshall, E. (2000). A renewed assault on an old and deadly foe. *Science* **290**: 428–430.

Matsubara, T., R. W. Beeman, H. Shike, N. J. Besansky, O. Mukabayire, S. Higgs, A. A. James, and J. C. Burns (1996). Pantropic retroviral vectors integrate and express in cells of the malaria mosquito, *Anopheles gambiae. Proc. Natl. Acad. Sci. USA* **93**: 6181–6185.

Matthews, P. R., M. B. Wang. P. M. Waterhouse, S. Thornton, S. J. Fieg, F. Gubler, and J. V. Jacobsen (2001). Marker gene elimination from transgenic barley, using co-transformation with adjacent "twin-T-DNAs" on a standard *Agrobacterium* transformation vector. *Mol. Breeding* **7**: 195–202.

McCarroll, L., M. G. Paton, S. H. P. P. Karunaratne, H. T. R. Jayasuryia, K. S. P. Kalpage, and J. Hemingway (2000). Insecticides and mosquito-borne disease. *Nature* **407**: 961–962.

McDermott, G.J., and M. A. Hoy (1997). Persistence, and containment of *Metaseiulus occidentalis* (Acari: Phytoseiidae) in Florida: risk assessment for possible releases of transgenic strains. *Florida Entomologist* **80**: 42–53.

McGrane, V., J. O. Carlson, B. R. Miller, and B. J. Beaty (1988). Microinjection of DNA into *Aedes triseriatus* ova and detection of integration. *Amer. J. Trop. Med. Hyg.* **39**: 502–510.

McKelvey, T. A., D. E. Lynn, D. Gundersen-Rindal, D. Guzo, D. A. Stoltz, K. P. Guthrie, P. B. Taylor, and E. M. Dougherty (1996). Transformation of gypsy moth (*Lymantria dispar*) cell lines by infection with *Glyptapanteles indiensis* polydnavirus. *Biochem. Biophys. Res. Comm.* **225**: 764–770.

Meredith, S. E. O., and A. A. James (1990). Biotechnology as applied to vectors and vector control. *Ann. Parasitol. Hum. Comp*. **65** (Suppl. I): 113–118.

Merrihew, R. V., W. C. Clay, J. P. Condreay, S. M. Witherspoon, W. S. Dallas, and T. A. Kost (2001). Chromosomal integration of transduced recombinant baculovirus DNA in mammalian cells. *J. Virol.* **75**: 903–909.

Mialhe, E., and L. H. Miller (1994). Biolistic techniques for transfection of mosquito embryos (*Anopheles gambiae*). *BioTechniques* **16**: 924.

Michel, K., A. Stamenova, A. C. Pinkerton, G. Franz, A. S. Robinson, A. Gariou-Papalexiou, A. Zacharopoulou, D. A. O'Brochta, and P. W. Atkinson (2001). *Hermes*-mediated germ-line transformation of the Mediterranean fruit fly *Ceratitis capitata*. *Insect Mol. Biol.* **10**: 155–162.

Miller, L. K. (1988). Baculoviruses as gene expression vectors. *Annu. Rev. Microbiol.* **42**: 177–199.

Miller, L. H. (1989). Strategies for malaria control: realities, magic, and science. In: *Biomedical Science and the Third World: Under the Volcano*. *Ann. N.Y. Acad. Sci.* **569**: 118–126.

Miller, L. H., R. K. Sakai, P. Romans, W. Gwadz, P. Kantoff, and H. G. Coon (1987). Stable integration and expression of a bacterial gene in the mosquito *Anopheles gambiae*. *Science* **237**: 779–781.

Moreira, L. A., M. J. Edwards, F. Adhami, N. Jasinskiene, A. A. James, and M. Jacobs-Lorena (2000). Robust gut-specific gene expression in transgenic *Aedes aegypti* mosquitoes. *Proc. Natl. Acad. Sci. USA* **97**: 10895–10898.

Mori, H. M. Yamao, H. Nakazawa, Y. Sugahara, N. Shirai, F. Matsubara, M. Sumida, and T. Imamura (1995). Transovarial transmission of a foreign gene in the silkworm, *Bombyx mori*, by *Autographa californica* nuclear polyhedrosis virus. *Bio/Technol.* **13**: 1005–1007.

Morris, A. C., P. Eggleston, and J. M. Crampton (1989). Genetic transformation of the mosquito *Aedes aegypti* by micro-injection of DNA. *Med. Vet. Entomol.* **3**: 1–7.

Morris, A. C., T. L. Schaub, and A. A. James (1991). FLP-mediated recombination in the vector mosquito, *Aedes aegypti. Nucleic Acids Res.* **19**: 5895–5900.

Morris, A. C., G. B. Pott, J. Chen, and A. A. James (1995). Transient expression of a promoter–reporter construct in differentiated adult salivary glands and embryos of the mosquito *Aedes aegypti*. *Am. J. Trop. Med. Hyg.* **52**: 456–460.

Nagaraju, J., T. Kanda, K. Yukuhiro, G. Chavancy, T. Tamura, and P. Couble (1996). Attempt at transgenesis of the silkworm (*Bombyx mori* L.) by egg-injection of foreign DNA. *Appl. Entomol. Zool.* **31**: 587–596.

Nikolaev, A. I., T. T. Tchkonia, C. A. Kafianieristavi, and V. X. Tarantul (1993). Preferential extrachromosomal localization of exogenous DNA in transgenic silkworm *Bombyx mori* L. *Mol. Gen. Genet.* **236**: 326–330.

Nikitin, A. G., and R. C. Woodruff (1995). Somatic movement of the *mariner* transposable element and lifespan of *Drosophila* species. *Mut. Res.* **338**: 43–49.

O'Brochta, D. A., and P. W. Atkinson (1997). Recent developments in transgenic insect technology. *Parasitol. Today* **13**: 99–104.

O'Brochta, D. A., and A. M. Handler (1988). Mobility of P elements in drosophilids and nondrosophilids. *Proc. Natl. Acad. Sci. USA* **85**: 6052–6056.

O'Brochta, D. A., W. D. Warren, K. J. Saville, and P. W. Atkinson (1996). *Hermes*, a functional non-drosophilid insect gene vector from *Musca domestica. Genetics* **142**: 907–914.

O'Brochta, D. A., P. W. Atkinson, and M. J. Lehane (2000). Transformation of *Stomoxys calcitrans* with a *Hermes* gene vector. *Insect Mol. Biol.* **9**: 531–538.

Olson, K. E. (2000). Sindbis virus expression systems in mosquitoes: background, methods, and applications, pp. 161–189. In: *Insect Transgenesis. Methods, and Applications*, A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Olson, K. E., S. Higgs, P. J. Gaines, A. M. Powers, B. S. Davis, K. I. Kamrud, J. O. Carlson, C. D. Blair, and B. J. Beaty (1996). Genetically engineered resistance to Dengue-2 virus transmission in mosquitoes. *Science* **272**: 884–886.

Olson, K. E., K. M. Myles, R. C. Seabaugh, S. Higgs, J. O. Carlson, and B. J. Beaty (2000). Development of a Sindbis virus expression system that efficiently expresses green fluorescent protein in midguts of *Aedes aegypti* following per os infection. *Insect Mol. Biol.* **9**: 57–65.

Omholt, S. W., S. Rishovd, A. Hagen, O. Elmholdt, B. Dalsgard, and S. Fromm (1995). Successful production of chimeric honeybee larvae. *J. Exp. Zool.* **272**: 410–412.

O'Neill, S. L., and T. L. Karr (1990). Bidirectional incompatibility between conspecific populations of *Drosophila simulans. Nature* **348**: 178–180.

Oosumi, T., W. R. Belknap, and B. Garlick (1995). *Mariner* transposons in humans. *Nature* **378**: 672.

Orr, W. C., and R. S. Sohal (1992). The effects of catalase gene overexpression on life span and resistance to oxidative stress in transgenic *Drosophila melanogaster. Arch. Biochem. Biophys.* **297**: 35–41.

Pal, R. (1974). WHO/ICMR programme of genetic control of mosquitos in India, pp. 73–95. In: *The Use of Genetics in Insect Control*. R. Pal, and M. J. Whitten, Eds. Elsevier, Amsterdam.

Pal, R., and M. J. Whitten, Eds. (1974). *The Use of Genetics in Insect Control*. Elsevier, Amsterdam.

Pal-Bhadra, M., U. Bhadra, and J. A. Birchler (1999). Cosuppression of nonhomologous transgenes in *Drosophila* involves mutually related endogenous sequences. *Cell* **99**: 35–46.

Pannuti, A., T. Kocacitak, and J. C. Lucchesi (2000). *Drosophila* as a model for the study of sex determination in anopheline and aedine mosquitoes, pp. 263–269, In: *Area-Wide Control of Fruit Flies and Other Insect Pests*. K. H. Tan, Ed. Penerbit Universiti Sains Malaysia, Penang.

Pelisson, A., L. Teysset, F. Chalvet, A. Kim, N. Prud'homme, C. Terzian, and A. Bucheton (1997). About the origin of retroviruses and the co-evolution of the *gypsy* retrovirus with the *Drosophila flamenco* host gene. *Genetica* **100**: 29–37.

Peloquin, J. J., S. T. Thibault, L. P. Schouest, and T. A. Miller (1997). Electromechanical microinjection of pink bollworm *Pectinophora gossypiella* embryos increases survival. *BioTechniques* **22**: 496–499.

Peloquin, J. J., S. T. Thibault, R. Staten, and T. A. Miller (2000a). Germ-line transformation of pink bollworm (Lepidoptera: Gelechiidae) mediated by the *piggyBac* transposable element. *Insect Mol. Biol.* **9**: 323–333.

Peloquin, J. J., L. Kuzina, C. R. Lauzon, and T. A. Miller (2000b). Transformation of internal extracellular bacteria isolated from *Rhagoletis completa* Cresson gut with enhanced green fluorescent protein. *Curr. Microbiol.* **40**: 367–371.

Peronnet, F., F. San Giorgioi, J. A. Lepesant, J. S. Deutsch, and G. Gonzy-Treboul (2000). Three-partner conversion induced by the P-element transposase in *Drosophila melanogaster. Mol. Gen. Genet.* **262**: 1123–1131.

Peters, I. D., D. E. Rancourt, P. L. Davies, and V. K. Walker (1993). Isolation, and characterization of an antifreeze protein precursor from transgenic *Drosophila*: evidence for partial processing. *Biochim. Biophys. Acta* **1171**: 247–254.

Peterson, K. R., C. H. Clegg, Q. Li, and G. Stamatoyannopoulos (1997). Production of transgenic mice with yeast artificial chromosomes. *Trends Genet.* **13**: 61–66.

Petrov, D. A., J. L. Schutzman, D. L. Hartl, and E. R. Lozovskaya (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis. Proc. Natl. Acad. Sci. USA* **93**: 8050–8054.

Pfeifer, T. A., and T. A. Grigliatti (1996). Future perspectives on insect pest management: engineering the pest. *J. Invert. Pathol.* **67**: 109–119.

Pinkerton, A. C., K. Michael, D. A. O'Brochta, and P. W. Atkinson (2000). Green fluorescent protein as a genetic marker in transgenic *Aedes aegypti. Insect Mol. Biol.* **9**: 1–10.

Plasterk, R. H. A. (1993). Molecular mechanisms of transposition and its control. *Cell* **74**: 781–786.

Plasterk, R. H. A., Z. Izsvak, and Z. Ivics (1999). Resident aliens. The Tc1/*mariner* superfamily of transposable elements. *Trends Genet.* **15**: 326–332.

Presnail, J. K., and M. A. Hoy (1992). Stable genetic transformation of a beneficial arthropod by microinjection. *Proc. Natl. Acad. Sci. USA* **89**: 7732–7726.

Presnail, J. K., and M. A. Hoy (1996). Maternal microinjection of the endoparasitoid *Cardiochiles diaphaniae* (Hymenoptera: Braconidae). *Ann. Entomol. Soc. Am.* **89**: 576–580.

Presnail, J.K., A. Jeyaprakash, J. Li, and M. A. Hoy (1997). Genetic analysis of four lines of *Metaseiulus occidentalis* (Nesbitt) (Acari: Phytoseiidae) transformed by maternal microinjection. *Ann. Entomol. Soc. Am.* **90**: 237–245.

Prout, T. (1994). Some evolutionary possibilities for a microbe that causes incompatibility in its host. *Evolution* **48**: 909–911.

Rancourt D. E., I. D. Peters, V. K. Walker, and P. L. Davies (1990). Wolffish antifreeze protein from transgenic *Drosophila*. *Bio/Technology* **8**: 453–457.

Rancourt, D. E., P. L. Davies, and V. K. Walker (1992). Differential translatability of antifreeze protein mRNAs in a transgenic host. *Biochim. Biophys. Acta* **1129**: 188–194.

Raymond, M., A. Callaghan, P. Fort, and N. Pasteur (1991). Worldwide migration of amplified insecticide resistance genes in mosquitoes. *Nature* **350**: 151–153.

Richards, F. F. (1993). An approach to reducing arthropod vector competence. Dispersion of insect-borne diseases could be modified by genetically altered symbionts. *Am. Soc. Microbiol. News* **59**: 509–514.

Robertson, H. M. (1993). The *mariner* transposable element is widespread in insects. *Nature* **362**: 241–245.

Robertson, H. M. (1995). The *Tc1-mariner* superfamily of transposons in animals. *J. Insect Physiol.* **41**: 99–105.

Robertson, H. M. (1997). Multiple *mariner* transposons in flatworms and hydras are related to those of insects. *J. Hered.* **88**: 195–201.

Robertson, H. M., and D. J. Lampe (1995). Distribution of transposable elements in arthropods. *Annu. Rev. Entomol.* **49**: 333–337.

Robertson, H. M., and R. Martos (1997). Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage. *Gene* **205**: 219–228.

Robinson, A. S., and G. Franz (2000). The application of transgenic insect technology in the sterile insect technique, pp. 307–318. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Robinson, K. O., H. J. Ferguson, S. Cobey, H. Vaessin, and B. H. Smith (2000). Sperm-mediated transformation of the honey bee, *Apis mellifera. Insect Mol. Biol.* **9**: 625–634.

Rong, Y. S., and K. G. Golic (2000). Site-specific recombination for the genetic manipulation of transgenic insects, pp. 53–75. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Rothenbuhler, W. C. (1979). Semidomesticated insects: honeybee breeding, pp. 84–92. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Rubin, G. M., and A. C. Spradling (1982). Genetic transformation of *Drosophila* with transposable element vectors. *Science* **218**: 348–353.

Ruesink J. L., I. M. Parker, M. J. Groom, and P. M. Kareiva (1995). Reducing the risks of nonindigenous species introductions: guilty until proven innocent. *Bio Science* **45**: 465–477.

Santamaria, P. (1986). Injecting eggs, pp. 159–173. In: Drosophila *A Practical Approach*. D. B. Roberts, Ed. IRL Press, Oxford.

Sarkar, A., and F. H. Collins (2000). Eye color genes for selection of transgenic insects, pp. 79–91. In: *Insect Transgenesis*. *Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Sarkar, A., C. J. Coates, S. Whyard, U. Willhoeft, P. W. Atkinson, and D. A. O'Brochta (1997a). The *Hermes* element from *Musca domestica* can transpose in four families of cyclorrhaphan flies. *Genetica* **99**: 15–29.

Sarkar, A., K. Yardley, P. W. Atkinson, A. A. James, and D. A. O'Brochta (1997b). Transposition of the *Hermes* element in embryos of the vector mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **27**: 359–363.

Schubbert, R., D. Renz, B. Schmitz, and W. Doerfler (1997). Foreign (M13) DNA ingested by mice reaches peripheral leukocytes, spleen, and liver via the intestinal wall mucosa and can be covalently linked to mouse DNA. *Proc. Natl. Acad. Sci. USA* **94**: 961–966.

Schubbert, R., U. Hohlweg, D. Renz, and W. Doerfler (1998). On the fate of orally ingested foreign DNA in mice: chromosomal association and placental transmission to the fetus. *Mol. Gen. Genet.* **259**: 569–576.

Schuler, M. L., H. A. Wood, R. R. Granados, and D. A. Hammer, Eds. (1994). *Baculovirus Expression Systems and Biopesticides*. Wiley-Liss, NY.

Seabaugh, R. C., K. E. Olson, S. Higgs, J. O. Carlson, and B. J. Beaty (1998). Development of a chimeric Sindbis virus with enhanced *per os* infection of *Aedes aegypti*. *Virology* **243**: 99–112.

Severson, D. W. (1994). Applications of molecular marker analysis to mosquito vector competence. *Parasitol. Today* **10**: 336–340.

Shamila, Y., and S. Mathavan (1998). Gene transfer in silkworm *Bombyx mori* via electroporation. *Biotechnol. Techn.* **12**: 507–510.

Sherman, A., A. Dawson, C. Mather, H. Gilhooley, Y. Li, R. Mitchell, D. Finnegan, and H. Sang (1998). Transposition of the *Drosophila* element *mariner* into the chicken germ line. *Nat. Biotechnol.* **16**: 1050–1053.

Shimizu, K., M. Kamba, H. Sonobe, T. Kanda, A. G. Klinakis, C. Savakis, and T. Tamura (2000). Extrachromosomal transposition of the transposable element *Minos* occurs in embryos of the silkworm *Bombyx mori. Insect Mol. Biol.* **9**: 277–281.

Sijen, T., and J. M. Kooter (2000). Post-transcriptional gene-silencing: RNAs on the attack or on the defense? *BioEssays* **22**: 520–531.

Silbermann, R., and M. Tatar (2000). Reproductive costs of heat shock protein in transgenic *Drosophila melanogaster*. *Evolution* **54**: 2038–2045.

Simpson, P. (1993). Flipping fruit-flies: a powerful new technique for generating *Drosophila* mosaics. *Trends Genet.* **9**: 227–228.

Sinkins, S. P., and S. L. O'Neill (2000). *Wolbachia* as a vehicle to modify insect populations, pp. 271–287. In: *Insect Transgenesis. Methods, and Applications*. A. M. Handler and A. J. James, Eds. CRC Press, Boca Raton, FL.

Spielman, A. (1994). Why entomological antimalaria research should not focus on transgenic mosquitoes. *Parasitol. Today* **10**: 374–376.

Spielman, A., J. C. Beier, and A. E. Kiszewski (2002). Ecological and community considerations in engineering arthropods to suppress vector-borne disease, pp. 325–329. In: *Genetically Engineered Organisms. Assessing Environmental and Human Health Effects*. D. K. Letourneau and B. E. Burrows, Eds. CRC Press, Boca Raton, FL.

Spradling, A. C., and G. M. Rubin (1982). Transposition of cloned P elements into *Drosophila* germline chromosomes. *Science* **218**: 341–347.

Stouthamer, R., R. F. Luck, and J. H. Werren (1992). Genetics of sex determination and the improvement of biological control using parasitoids. *Environ. Entomol.* **21**: 427–435.

Stowell, S. L., R. E. Sinden, and J. M. Crampton (1998). Expressing immunogenic parasite proteins in the mosquito: a model system to examine the possibility of using haematophagous insects to deliver vaccines. Abstract 216, p. 409. In: Keystone Symposium: *Toward the Genetic Manipulation of Insects*, Taos, NM.

Sundararajan, P., P. W. Atkinson, and D. A. O'Brochta (1999). Transposable element interactions in insects: crossmobilization of *hobo* and *Hermes. Insect Mol. Biol.* **8**: 359–368.

Tamura, T., T. Thibert, C. Royer, T. Kanda, A. Eappen, M. Kamba, N. Komoto, J. L. Thomas, B. Mauchamp, G. Chavancy, P. Shirk, M. Fraser, J. C. Prudhomme, and P. Couble (2000). A *piggyBac* element-derived vector efficiently promotes germline transformation in the silkworm *Bombyx mori* L. *Nat. Biotechnol.* **18**: 81–84.

Tan, K. H., Ed. (2000). Area-*Wide Control of Fruit Flies and Other Insect Pests*. Penerbit Universiti Sains Malaysia, Penang.

Tatar, M., A. Kopelman, D. Epstein, M. P. Tu, C. M. Yin, and R. S. Garofalo (2001). A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**: 107–110.

Terzian, C., A. Pelisson, and A. Bucheton (2000). Retrotransposons and retroviruses in insect genomes, pp. 191–215. In: *Insect Transgenesis. Methods and Applications.* A. M. Handler and A. A. James, Eds. CRC Press, Boca Raton, FL.

Thomas, D. D., C.A. Donnelly, R. J. Wood, and L. S. Alphey (2000). Insect population control using a dominant, repressible, lethal genetic system. *Science* **298**: 2474–2476.

Tiedje, J. M., R. K. Colwell, Y. L. Grossman, R. E. Hodson, R. E. Lenski, R. M. Mack, and P. J. Regal (1989). The planned introduction of genetically engineered organisms: ecological considerations and recommendations. *Ecology* **70**: 298–315.

Tower, J. (2000). Transgenic methods for increasing *Drosophila* life span. *Mech. Aging Dev.* **118**: 4–14.

Tsitrone, A., S. Charles, and C. Biemont (1999). Dynamics of transposable elements under the selection model. *Genet. Res. Camb.* **74**: 159–164.

Turelli, M., and A. A. Hoffmann (1991). Rapid spread of an inherited incompatibility factor in California *Drosophila. Nature* **353**: 440–442.

Turelli, M., A. A. Hoffmann, and S. W. McKechnie (1992). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics* **132**: 713–723.

U. S. Department of Agriculture (1991). Part III. Proposed guidelines for research involving the planned introduction into the environment of organisms with deliberately modified hereditary traits; Notice. Federal register Vol. **56**(22), Friday, February 1, 1991, pp. 4134–4151.

U. S. Department of Agriculture, Animal and Plant Health Inspection Service (2002). The regulation of transgenic arthropods. http://www.aphis.usda.gov:80/bbep/bp/arthropod/#tgenadoc

Vizioli, J., A. M. Richman, S. Uttenweiler-Joseph, C. Blass, and P. Bulet (2001). The defensin peptide of the malaria vector mosquito *Anopheles gambiae*: antimicrobial activities and expression in adult mosquitoes. *Insect Biochem. Mol. Biol.* **31**: 241–248.

Vollrath, F., and D. P. Knight (2001). Liquid crystalline spinning of spider silk. *Nature* **410**: 541–548.

Walker, V. K. (1989). Gene transfer in insects. *Adv. Cell Culture* **7**: 87–124.

Walker, V. K., D. E. Rancourt, and B. P. Duncker (1995). The transfer of fish antifreeze genes to *Drosophila:* a model for the generation of transgenic beneficial insects. *Proc. Entomol. Soc. Ontario* **126**: 3–13.

Wang, W., L. Swevers, and K. Iatrou (2000). *Mariner (Mos1)* transposase and genomic integration of foreign gene sequences in *Bombyx mori* cells. *Insect Mol. Biol.* **9**: 145–155.

Watanabe, K., and M. Sato (1998). Plasmid-mediated gene transfer between insect-resident bacteria, *Enterobacter cloacae,* and plant-epiphytic bacteria, *Erwinia herbicola*, in guts of silkworm larvae. *Curr. Microbiol.* **37**: 352–355.

Watanabe, K., W. Hara, and M. Sato (1998). Evidence for growth of strains of the plant epiphytic bacterium *Erwinia herbicola* and transconjugation among the bacterial strains in guts of the silkworm *Bombyx mori. J. Invert. Pathol.* **72**: 104–111.

Webb, B. A. (2000). Polydnaviruses, and insect transgenic research, pp. 203–215. In: *Insect Transgenesis. Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

White, L. D., C. J. Coates, P. W. Atkinson, and D. A. O'Brochta (1996). An eye color gene for the detection of transgenic non-drosophilid insects. *Insect Biochem. Mol. Biol.* **26**: 641–644.

Whitten, M. J. (1979). The use of genetically selected strains for pest replacement or suppression, pp. 31–40. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Whitten, M. J. (1985). The conceptual basis for genetic control, pp. 465–527. In: *Comprehensive Insect Physiology Biochemistry and Pharmacology*. G. A. Kerkut, and L. I. Gilbert, Eds. Pergamon Press, Oxford.

Whitten, M. J., and M. A. Hoy (1999). Genetic improvement and other genetic considerations for improving the efficacy and success rate of biological control, pp. 271–296. In: *Principles and Applications of Biological Control*. T. S. Bellows, and T. W. Fisher, Eds. Academic Press, San Diego.

Williamson, M. (1992). Environmental risks from the release of genetically modified organisms (GMOs)—the need for molecular ecology. *Mol. Ecol.* **1**: 3–8.

Witte, W. (1998). Medical consequences of antibiotic use in agriculture. *Science* **279**: 996–997.

Wolffe, A. P. (1997). Transcription control: repressed repeats express themselves. *Curr. Biol.* **7**: R796–R798.

Wood, S. N., and M. B. Thomas (1999). Super-sensitivity to structure in biological models. *Proc. R. Soc. Lond. B* **266**: 565–570.

Wright, J. W., and R. Pal, Eds. (1967). *Genetics of Insect Vectors of Disease*. Elsevier, Amsterdam.

Wu, C., S. Asakawa, N. Shimizu, S. Kawasaki, and Y. Yasukochi (1999). Construction, and characterization of bacterial artificial chromosome libraries from the silkworm *Bombyx mori*. *Mol. Gen. Genet.* **261**: 698–706.

Wyss, J. H. (2000). Screw-worm eradication in the Americas—Overview, pp. 79–86. In: *Area-Wide Control of Fruit Flies and Other Insect Pests*. K. H. Tan, Ed. Penerbit Universiti Sains Malaysia, Penang.

Yamao, M., N. Katayama, H. Nakazawa, M. Yamakawa, Y. Hayashi, S. Hara, K. Kamei, and H. Mori (1999). Gene targeting in the silkworm by the use of a baculovirus. *Genes Dev.* **13**: 511–516.

Yan, J., Q. Cheng, C. B. Li, and S. Aksoy (2001). Molecular characterization of two serine proteases expressed in gut tissue of the African trypanosome vector, *Glossina morsitans morsitans*. *Insect Mol. Biol.* **10**: 47–56.

Yen, J. H., and A. R. Barr (1974). Incompatibility in *Culex pipiens*, pp. 97–118. In: *The Use of Genetics In Insect Control*. R. Pal, and M. J. Whitten, Eds. Elsevier, Amsterdam.

Yoder, J. I., and A. P. Goldsbrough (1994). Transformation systems for generating marker-free transgenic plants. *Bio/Technol.* **12**: 263–267.

Yokoyama, T. (1979). Silkworm selection and hybridization, pp. 71–83. In: *Genetics in Relation to Insect Management*. M. A. Hoy, and J. J. McKelvey, Jr., Eds. Rockefeller Foundation Press, New York.

Yoshida, S., D. Ioka, H. Matsuoka, H. Endo, and A. Ishii (2001). Bacteria expressing single-chain immunotoxin inhibit malaria parasite development in mosquitoes. *Mol. Biochem. Parasitol.* **113**: 89–96.

Yoshiyama, M., H. Honda, and K. Kimura (2000a). Successful transformation of the housefly, *Musca domestica* (Diptera: Muscidae) with the transposable element, *mariner*. *Appl. Entomol. Zool.* **35**: 321–325.

Yoshiyama, M. K. Kimura, and H. Honda (2000b). Technique for microinjection into eggs of the housefly, *Musca domestica* (Diptera: Muscidae). *Appl. Entomol. Zool.* **35**: 27–30.

Young, O. P., S. P. Ingebritsen, and A. S. Foudin (2000). Regulation of transgenic arthropods and other invertebrates in the United States, pp. 369–379. In: *Insect Transgenesis. Methods and Applications.* A. M. Handler, and A. A. James, Eds. CRC Press, Boca Raton, FL.

Yu, Z., J. D. Podgwaite, and H. A. Wood (1992). Genetic engineering of a *Lymantria dispar* nuclear polyhedrosis virus for expression of foreign genes. *J. Gen. Virol.* **73**: 1509–1514.

Zalokar, M. (1981). A method for injection and transplantation of nuclei and cells in *Drosophila* eggs. *Experientia* **37**: 1354–1356.

Zhao, Y. G., and P. Eggleston (1998). Stable transformation of an *Anopheles gambiae* cell line mediated by the *Hermes* mobile genetic element. *Insect Biochem. Mol. Biol.* **28**: 213–219.

Zhao, Y. G., and P. Eggleston (1999). Comparative analysis of promoters for transient gene expression in cultured mosquito cells. *Insect Mol. Biol.* **8**: 31–38.

Zheng, L. (1997). Molecular approaches to mosquito parasite interactions. *Arch. Insect Biochem. Physiol.* **34**: 1–18.

Zheng, L. B., A. J. Cornel, R. Wang, H. Erfle, H. Voss, W. Ansorge, F. C. Kafatos, and F. H. Collins (1997). Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium cynomolgi* B. *Science* **276**: 425–428.

Zwiebel, L. J., G. Saccone, A. Zacharopoulou, N. J. Besansky, G. Favia, F. H. Collins, C. Louis, and F. C. Kafatos (1995). The *white* gene of *Ceratitis capitata:* a phenotypic marker for germline transformation. *Science* **270**: 2005–2008.

Some Relevant Web Sites

Information Systems for Biotechnology, a site that provides information on agricultural biotechnology, is at: www.nbiap.vt.edu

Malaria International Foundation provides extensive information on malaria at: www.malaria.org

Pew Initiative on Food and Biotechnology: http://pewagbiotech.org/about/

The Mosquito Genome Database Server is at: www.malaria.org

Transformation of Mosquito Vectors Action Plan: http://klab.agsci.colostate.edu/mfnet/mftacp.html

U. S. Department of Agriculture, Animal and Plant Health Inspection Service. The regulation of transgenic arthropods is at: www.aphis.usda.gov:80/bbep/bp/arthropod/#tgenadoc

GLOSSARY

A-DNA Right-handed helical form of DNA found in fibers at 75% RH in presence of sodium, potassium or cesium. The bases are tilted with regard to the helical axis and there are more base pairs per turn. The A-form may be very similar to the conformation adopted by DNA-RNA hybrids or by RNA-RNA double-stranded regions.

acentric A chromosome, or chromosome fragment, that lacks a centromere.

actin A protein that is the major constituent of the microfilaments of cells and in muscle. All actins studied are similar in size and amino acid sequence, suggesting they have evolved from a single gene.

action potential A rapid change in the polarity of the membrane of a neuron that facilitates the interaction and transmission of impulses.

additive genes Genes that interact but do not show dominance or epistasis.

adenine A purine and one of the nitrogenous bases found in DNA and RNA.

adenosine triphosphate See ATP.

agarose A polysaccharide gum obtained from agar, which is obtained from certain seaweeds, used as a gel medium in electrophoresis; used to separate DNA molecules on the basis of their molecular weight.

allele One of two or more alternative forms of a gene at a particular locus. If more than two alleles exist, the locus is said to exhibit multiple allelism.

allopatry Populations that are physically or geographically isolated eventually may change sufficiently through natural selection or drift that pre- or postmating isolation mechanisms develop, leading eventually to speciation.

allozyme Allozymes are a subset of isozymes. Allozymes are variants of enzymes representing different allelic alternatives of the same locus.

alternative splicing Gene regulation by means of alternative splicing of exons to produce different amounts of protein or even different proteins.

amino acid One of the monomeric units that polymerize to make a protein molecule.

aminoacyl tRNA synthetase Enzymes that catalyze the attachment of each amino acid to the appropriate transfer RNA molecule. A tRNA molecule carrying its amino acid is called a charged tRNA.

amplification The production of additional copies of a chromosomal sequence, found as either intrachromosomal or extrachromosomal DNA.

anchored PCR A modification of PCR that allows amplification in situations in which only one sequence is known that is suitable for a primer (rather than two). The procedure involves synthesis of cDNA with the known primer from mRNA. A polyG tail is added to the cDNA. The second primer is developed by synthesizing a primer with a polyC sequence, which allows amplification of a second DNA strand that is complementary to the cDNA.

Subsequent cycles yield amplified DNA from both strands.

aneuploid A condition in which the chromosome number of an organism is not an exact multiple of the typical haploid set for the species.

angstrom Abbreviated as Å; one hundred-millionth of a centimeter, or 0.1 nm.

anneal The process by which the complementary base pairs in the strands of DNA combine.

anticodon The triplet of nucleotides in a transfer RNA molecule that is complementary to and base pairs with a codon in a messenger RNA.

antiparallel The DNA strands are parallel but point in opposite directions.

apomorphic A character that is derived and not ancestral.

apoptosis Programmed cell death, is a series of programmed steps that cause a cell to die via "self digestion" without rupturing and releasing intracellular contents into the surrounding environment.

arbitrarily primed PCR (AP-PCR) Does not require a particular set of primers; rather it uses a single primer chosen without regard to the sequence to be fingerprinted. By using a single primer and two cycles of low-stringency PCR followed by many cycles of high-stringency PCR, discrete and reproducible products characteristic of specific genomes are produced. As originally described, the primers used are 20-bp sequencing primers.

arrhenogenic A sex-determining system in which females produce male progeny only. Found in the blowfly *Chrysomya rufifacies* (Calliphoridae).

arrhenotoky A form of parthenogenesis in which an unfertilized egg develops into a male by parthenogenesis and a fertilized egg develops into the female. Arrhenotoky is found in many Hymenoptera.

associative learning The process of learning through the formation of associations between stimuli and responses.

asymmetric PCR Single-stranded DNA produced by providing an excess of primer for one of the two DNA strands. Asymmetric primer ratios are typically 50:1 to 100:1. Single-stranded DNA produced can be sequenced directly without cloning.

ATP Adenosine triphosphate is the primary molecule for storing chemical energy in a cell.

autecology The ecology of an individual organism or species.

autoradiography A method for detecting radioactively labeled molecules through exposure of an X-ray-sensitive photographic film.

autoregulatory control Regulation of the synthesis of a gene product by the product itself. In some systems, excess gene product behaves as a repressor and binds to the operator of its own structural gene.

autosomes All chromosomes except the sex chromosomes. Each diploid cell has two copies of each autosome.

B chromosomes B chromosomes are nonvital supernumerary chromosomes found in many plant and animal species, thought to be derived from one of the normal chromosomes, and are often transmitted at higher rates than expected, thus exhibiting "drive." The PSR condition (paternal sex ratio) of the parasitic wasp *Nasonia vitripennis* is an example of a B chromosome.

B-DNA A helical form of DNA formed by adding water to dehydrated A-DNA. B-DNA is the form of DNA of which Watson and Crick constructed their model in 1953. It is found in very high relative humidity. This form is thought to prevail in the living cell.

back mutation Mutations that occur to reverse a point mutation to the original condition.

bacterial conjugation A temporary union between two bacteria, in which genetic material is exchanged; DNA from the "male" cell transfers all or part of its chromosomes into the recipient "female."

bacteriophage A virus whose host is a bacterium. See lambda for description of λ phage.

baculovirus An insect-pathogenic virus with a circular double-stranded DNA genome and rod-shaped enveloped virion, found primarily in lepidopterans. These viruses have been engineered for two purposes: 1) as expression

vectors to produce large quantities of proteins or 2) as biological pesticides.

baculovirus expression vectors Vectors used to infect certain types of insect cells to produce particular proteins.

balancer chromosomes Balancer chromosomes initially were developed by H. J. Muller as a method for maintaining lethal mutations in laboratory stocks without continuous selection. Balancer *Drosophila* stocks contain several recessive visible mutations, one or more inversions and transpositions on a specific chromosome. These mutations have been induced to suppress crossing over. Balancing chromosomes may also have a clearly visible dominant mutation so that heterozygous files can be identified easily.

base pair (bp) Two nucleotides that are in different strands of nucleic acid and whose bases pair by hydrogen bonding. In DNA, adenine pairs with thymine and guanine pairs with cytosine.

bioinformatics Researchers in bioinformatics develop computer software applications that can store, compare, and analyze the very large quantities of DNA sequence data generated by the new genome technologies. New bioinformatics tools can sift through a mass of raw data, finding and extracting relevant information and their relationships. See also functional genomics, genomics, and structural genomics.

biotechnology The manipulation of organisms to provide desirable products for human use. It has broader meanings, as well, including all parts of an industry that creates, develops, and markets a variety of products through the molecular manipulation of organisms or using knowledge pertaining to organisms.

BLAST Basic Local Alignment Search Tool (BLAST) is a computer program widely used to search large databases of DNA or amino acid sequences, providing sequences that have regions of similarity to the sequence(s) of interest provided by the user.

blastoderm The layer of cells in an insect embryo that surrounds an internal yolk mass. The cellular blastoderm develops from a syncytium by surrounding the cleavage nuclei with membranes derived from the enfolding of the surrounding membrane.

blunt end An end of a DNA molecule, at which both strands terminate at the same nucleotide position with no extension of one of the strands.

bootstrapping A statistical method based on repeated random sampling with replacement from an original sample to provide a collection of new estimates of a parameter, from which confidence limits can be calculated.

C-banding Dark bands on chromosomes produced by strong alkaline treatment at high temperature followed by incubation in sodium citrate, followed by Giemsa staining. C-bands correspond to regions of constitutive heterochromatin.

C value paradox C stands for "constant" or "characteristic" and denotes the fact that the DNA content (size) of the haploid genome is fairly constant within a species. C values vary widely among species. Size is usually measured in picograms of DNA.

capping The modification of the $5'$ end of the pre-mRNA in eukaryotes when a GTP is added to the molecule via an unusual $5'$–$5'$ triphosphate bond. Capping is necessary for the ribosome to bind with the mRNA to begin protein synthesis.

carbohydrate A large class of carbon–hydrogen–oxygen compounds, including simple sugars (monosaccharides) such as glucose. Glucose is the major fuel for most organisms and is the basic building block of polysaccharides such as starch and cellulose.

cDNA The DNA copy of a eukaryotic messenger RNA molecule, produced *in vitro* by enzymatic synthesis and used for producing cDNA libraries or probes for isolating genes in genomic libraries.

cDNA library A collection of clones containing dsDNA that is complementary to the mRNA. Such clones will lack introns and regulatory regions of eukaryotic genes. Once cDNA molecules are transcribed, they are inserted into a vector and amplified in *E. coli*. Genes that are inactive will not be represented in a

cDNA library, nor will noncoding regions of the genome.

cell The fundamental unit of life; each multicelled organisms is composed of cells; cells may be organized into organs that are relatively autonomous but cooperate in the functioning of the organism.

cell-autonomous determination The establishment of a developmental pathway within a particular cell. Determination is not influenced by substances diffusing from elsewhere.

cell cycle The sequence of events between one cell mitotic division and another in a eukaryotic cell. Mitosis (M phase) is followed by a growth (G_1) phase, then by DNA synthesis (S phase), then by another growth (G_2) phase, and then by another mitosis.

Central Dogma The Central Dogma was proposed by F. Crick in 1958. It states that the genetic information is contained in DNA, which is transcribed into RNA, which is translated into polypeptides. The transfer of information was proposed to be unidirectional from DNA to polypeptides: polypeptides are unable to directly synthesize RNA, and RNA is unable to directly synthesize DNA. The Central Dogma was modified in 1970 when RNA viruses were found to transfer information from RNA to DNA.

centromere A region of a chromosome to which spindle fibers attach during mitosis and meiosis. The position of the centromere determines whether the chromosome will appear as a rod, a J, or a V during migration of the chromosome to the poles in anaphase. In some insects, the spindle fibers attach throughout the length of the chromosome and such chromosomes are called holocentric. Centromeres are usually bordered by heterochromatin containing repetitive DNA.

chain-terminating method of DNA sequencing See dideoxy sequencing.

channels See ion channels.

chaperones Protein molecules that assist with correct protein folding as the protein emerges from the cell's ribosome. Heat shock protein 70, heat shock protein 40, and chaperonins are examples.

chelating agent A molecule capable of binding metal atoms; one example is EDTA, which binds Mg^{2+}.

chemotaxis The movement of a cell or organisms toward or away from a chemical substance.

chiasmata Chiasmata occur during prophase I of meiosis and represent points where crossing over, or exchange of genetic information, between nonsister chromatids occurred. When the synapsed chromosomes begin to separate in late prophase I, they are held together by these connections between the chromatids of homologous chromosomes.

chimeric DNA Recombinant DNA containing DNA from two different species.

chitin A water-insoluble polysaccharide that forms the exoskeletons of arthropods and crustaceans.

chitinase An enzyme that degrades chitin.

chorion A complex structure covering the insect egg.

chromatids Chromosome components that have duplicated during interphase become visible during the prophase stage of mitosis. Chromatids are held together at the centromere.

chromomere A region on a chromosome consisting of densely packed chromatid fibers that produce a dark band on polytene chromosomes.

chromosomes Units of the genome with many genes, consisting of histone proteins and a very long DNA molecule; found in the nucleus of every eukaryote.

chromosome imprinting The mechanism involved in chromosomal imprinting, or labeling of DNA, may involve methylation of DNA in many organisms. Imprinting is a reversible, differential marking of genes or chromosomes that is determined by the sex of the parent from whom the genetic material is inherited.

chromosome puffs A localized swelling of a region of a polytene chromosome due to synthesis of DNA or RNA. Puffing is readily seen in polytene salivary gland chromosomes of dipteran insects.

chromosome walking A molecular genetic technique that allows a series of overlapping fragments of DNA to be ordered. The technique is used to isolate a gene of interest for which no probe is available but that is known to be linked to a gene which has been identified and cloned. The marker gene is used to screen a genomic library. All fragments containing the known cloned gene are selected and sequenced, the fragments are then aligned and those cloned segments farthest from the marker gene in both directions are subcloned for the next step, and so on. The subclones are used as probes to screen the genomic library to identify new clones containing DNA with overlapping sequences. As the process is repeated, the nucleotide sequences of areas farther and farther away from the marker gene are identified, and eventually the gene of interest will be found.

circadian clock Changes in biological or metabolic functions that show periodic peaks or lows of activity based on or approximating a 24-h cycle.

circadian rhythms Biological rhythms with periods of ∼24 h; rhythmicity is endogenous and self-sustaining, continuing under constant environmental conditions for a period.

clade An evolutionary lineage derived from a single stem species. A branch of a cladogram.

cladistic systematics Systematics that use only shared and derived characters as a basis of constructing classifications. The rate or amount of change subsequent to splitting of phyletic lines is not considered. All taxa must arise from a common ancestral species.

cladogenic speciation Branching evolution of new species.

cladogram A term used two ways by different authors. Either a dendrogram (tree) produced using the principle of parsimony, or a tree that depicts inferred historical relationships between organisms.

clock The entire circadian system, including the central oscillator and the input and output pathways. A clock can be entrained to the environmental day–night cycle and is used to track the passage of time by an organism.

clone A population of identical cells often containing identical recombinant DNA molecules. Also a group of organisms produced from one individual cell through asexual processes. The offspring are identical. The word may be used as either a noun or a verb.

cloning vector A DNA molecule capable of replicating in a host organism; a gene is often inserted into it to construct a recombinant DNA molecule, and the vector is then used to amplify (clone) the recombinant DNA.

cluster analysis A method of hierarchically grouping taxa or sequences on the basis of similarity or minimum distance. UPGMA is an unweighted pair group method using the arithmetic average. WPGMA is the weighted pair group method using the arithmetic average.

coding strand The strand of the DNA molecule that carries the biological information of a gene and that is transcribed by RNA polymerase into pre-mRNA.

codominant Alleles whose gene produces are both manifested in the heterozygote.

codon A triplet of nucleotides that codes for a single amino acid.

coefficient of gene differentiation Interpopulation diversity using allozyme data are usually measured using the coefficient of gene differentiation (G_{ST}). G_{ST} is derived by estimating the average similarity within and between populations. G_{ST} is an extension of Wright's correlation (F_{ST}) between two gametes drawn at random from each subpopulation. The coefficient of differentiation is $G_{ST} = (H_T - H_S)/H_T$, where H_S is the average gene diversity within populations and H_T is the interpopulation gene diversity.

colony Growth of a group of microorganisms derived from a single cell. After growth on appropriate media, the population is visible without magnification.

colony hybridization The use of *in situ* hybridization to identify bacterial colonies carrying inserted DNA that is homologous with a particular sequence (the probe).

competent cells Bacterial cells in a state in which exogenous DNA molecules can bind

and be internalized, thereby allowing transformation.

complementary base pairing Nucleotide sequences are able to base pair; A and T are complementary; 5′-ATGC-3′ is complementary to 5′-GCAT-3′.

complementary DNA (cDNA) An ss DNA that is complementary to a strand of RNA. The DNA is synthesized by an enzyme called reverse transcriptase. It is a DNA copy of the messenger RNA. A second step makes it ds.

concatemers The linking of multiple subunits into a tandem series or chain results in structures called concatemers.

concerted evolution Maintenance of homogeneity of nucleotide sequences among members of a gene family in a species even though the sequences change over time. Members of a gene family evolving in a nonindependent fashion.

conditional lethal A mutation that may be lethal only under certain environmental conditions.

conditioned stimulus A stimulus that evokes a response that was previously elicited by an unconditioned stimulus.

constitutive enzymes Enzymes that are part of the basic permanent machinery of the cell. They are formed consistently in constant amounts regardless of the metabolic state of the organism.

constitutive heterochromatin Regions of the chromosome containing mostly highly repeated, noncoding DNA; usually near the telomeres and centromeres.

contig Segments of DNA that partially overlap in their sequence are called contigs.

convergent evolution The evolution of unrelated species resulting in structures with a superficial resemblance.

conversion of transposable elements Conversion of disabled transposable elements into active elements sometimes can occur through a DNA repair process.

copy number The number of plasmids in a cell; the number of genes, transposons, or repetitive elements in a genome.

core DNA The DNA in the core nucleosome that is wrapped around the histone octamer. The core nucleosome is connected to others by linker DNA.

***cos* sites** The *cos* sites are cohesive end sites or nucleotide sequences that are recognized when a phage DNA molecule is being packaged into its protein coat.

cosmid Engineered vectors used to clone large segments of exogenous DNA, derived by inserting *cos* sites from phage λ into a plasmid. The resulting hybrid molecule can be packaged in the protein coat of a phage.

crossing over The reciprocal exchange of polynucleotides between homologous chromosomes during meiosis.

cytochrome The complex protein respiratory enzymes occurring within plant and animal cells in the mitochondria, where they function as electron carriers in biological oxidation.

cytoplasm The components of the cell *not* including the nucleus.

cytoplasmic incompatibility Reproductive incompatibility between two populations caused by factors that are present in the cytoplasm. Often associated with microorganisms.

cytoplasmic inheritance See maternally inherited.

cytoplasmic sex-ratio distorters Cytoplasmic genes that manipulate the sex ratio of their host to promote their own spread. Microbes (*Wolbachia*, spiroplasmas, viruses) often are transovarially and transstadially transmitted that can alter the sex ratios of insects and mites.

cytosine A pyrimidine, one of the bases in DNA and RNA.

cytosol The fluid portion of the cytoplasm, excluding the organelles in a cell.

dalton A unit of mass very nearly equal to that of a hydrogen atom.

degeneracy Refers to the genetic code and the fact that most amino acids are coded for by more than one triplet codon.

degenerate codons Two or more codons that code for the same amino acid.

degenerate primers Degenerate primers can be used for the PCR when a limited portion of a protein sequence is known for a gene, but the DNA sequence is not known.

deletion The loss of a portion of the genetic material from a chromosome. The size can vary from one nucleotide to sections containing many genes.

denaturation Breakdown of secondary and higher levels of structure of proteins or nucleic acids by chemical or physical means.

denatured DNA DNA that has been converted from double- to single-stranded form by a process such as heating.

dendrogram A branched diagram that represents the evolutionary history of a group of organisms.

density gradient centrifugation Separation of molecules and particles on the basis of buoyant density, often by centrifugation in a concentrated sucrose or cesium chloride solution.

deoxyribonuclease An enzyme that breaks a DNA polynucleotide by cleaving phosphodiester bonds.

deoxyribonucleic acid (DNA) The genetic information.

deuterotoky A form of parthenogenesis in which unfertilized eggs can develop into either males or females.

dicentric A chromosome or chromatid with two centromeres.

dideoxy sequencing Developed by F. Sanger and A. R. Coulson in 1975, and known as the "plus and minus" or "primed synthesis" method of DNA sequencing. DNA is synthesized *in vitro* so that it is radioactively labeled and the reaction terminates precisely at the position corresponding to a specific base. After denaturation, fragments of different lengths are separated by electrophoresis and identified by autoradiography. In the "plus" protocol, only one kind of dNTP is available for elongation of the ^{32}P-labeled primer. In the "minus" protocol, one of the four dNTPs is missing, or specific terminator base analogues are used.

diploid Having two copies of each chromosome.

direct repeats When a transposable element is inserted into a host genome, a small segment, typically 4 to 12 bp, of the host DNA is duplicated at the insertion site. The duplicated repeats are in the same orientation and are called direct repeats.

discontinuous gene A gene in which the genetic information is separated into two or more different exons by an intervening sequence (intron) which typically is noncoding. Most eukaryotic genes are discontinuous.

discrete character A character that is countable.

distance A measure of the difference between two objects.

distance estimates A phrase used to emphasize the fact that evolutionary history is inferred from experimental or sequence data, and distance is thus an estimate.

DNA Deoxyribonucleic acid, the genetic molecule.

DNA arrays DNA arrays work by hybridization of labeled RNA or DNA in solution to DNA molecules attached at specific locations on a surface. The hybridization of a sample to an array involves each molecule's "search" for a matching partner on the matrix with the eventual pairings of molecules on the surface determined by base complementarity. Original arrays involved DNA from cDNA, genomic DNA, or plasmid libraries spotted on a porous membrane with the hybridized material labeled with a radioactive group. Now, glass is often used as a substrate and fluorescence for detection. New technologies allow synthesizing or depositing nucleic acids on glass slides at very high densities, which means the nucleic acid arrays have become miniaturized, leading to increased efficiency and information content.

DNA binding protein Proteins such as histones or RNA polymerase that attach to DNA as part of their function.

DNA–DNA hybridization A method for determining the degree of sequence similarity between DNA strands from two different organisms by the formation of heteroduplex molecules.

DNA fingerprinting See fingerprinting.

DNA ligase An enzyme that repairs single-stranded discontinuities in double-stranded DNA. DNA ligases also are used in constructing recombinant DNA molecules.

DNA methylation See methylation.

DNA polymerase An enzyme that catalyzes the formation of DNA from dNTPs, using single-stranded DNA as a template. Three different DNA polymerases (I, II, and III) have been isolated from *E. coli.* Eukaryotes contain different DNA polymerases, found in the nucleus, cytoplasm, or mitochondria, that are involved in DNA replication, repair, and recombination.

DNA polymerase I The enzyme in *E. coli* that completes synthesis of individual Okazaki fragments during DNA replication.

DNA polymerase III The enzyme that primarily functions in DNA replication of *E. coli*.

DNA probe Also called a gene probe or genetic probe. Short, specific (complementary to the desired DNA sequence), artificially produced segments of labeled DNA are used to combine with and detect the presence of a specific gene or DNA sequence within the chromosome. The presence of this labeled probe usually is detected visually.

DNA sequencing Determining the order of nucleotides in a DNA molecule.

DNA topoisomerase An enzyme that introduces or removes turns from the double helix by transiently breaking one or both of the strands.

DNA vaccines When a strand of DNA that has been extracted from or derived from a pathogen is injected into tissues in the host organism, the tissues may take up the naked DNA and express some of the cell-surface proteins of the pathogen. If the host's immune system mounts an immune response to those proteins (and thus to the pathogen), the injected naked genes are referred to as DNA vaccines.

DNase Deoxyribonuclease, an enzyme that degrades DNA.

dominant A gene is dominant when it produces the same phenotype whether it is heterozygous or homozygous. The trait is expressed even if only one copy of the gene is present in the genome.

dosage compensation A mechanism that compensates for the dosage of genes carried on the X chromosome in XX and XY organisms. In mammals, one or more of the X chromosomes is inactivated. In *Drosophila* males the genes on the Y chromosome are hypertranscribed.

double helix The base-paired structure consisting of two polynucleotides in the natural form of DNA.

downstream Toward the 3′ end of a DNA molecule.

driver Unlabeled DNA used in DNA–DNA hybridization.

ds DNA Double-stranded DNA.

ecdysone A steroid hormone found in insects that initiates and coordinates the molting process and the sequential expression of stage-specific genes.

EDTA ETDA, "ethylenediaminetetraacetic acid," a chelating agent, is able to react with metallic ions, even in minute amounts, and form a stable, inert, water-soluble complex.

electrophoresis The separation of molecules in an electric field. Electrophoresis can be used to separate proteins or DNA molecules.

electroporation A process used to introduce DNA into the genome of an organism. Electroporation uses a brief direct-current electrical pulse to open "micropores" in the surfaces of cells suspended in water containing DNA sequences. After the DNA enters the cell via the micropores, the electrical pulse ends and the pores close. The cell then can incorporate some of the DNA into its genome.

enantiomers Compounds showing mirror-image isomerism.

endonuclease An enzyme which degrades nucleic acid molecules by cleaving phosphodiester bonds internally.

endoplasmic reticulum (ER) A system of sacs (cisternae) in the cytoplasm of eukaryotic cells in which the ER is continuous with the plasma membrane and the outer membrane of the nuclear envelope. If the outer surfaces of the ER membranes are coated with ribosomes,

the ER is "rough-surfaced"; otherwise it is called smooth-surfaced.

endopolyploidy The occurrence in a diploid individual of cells containing 4-, 8-, 16-, 32-fold, etc., amounts of DNA in their nuclei. Nurse cells of ovaries are often endopolyploid.

endosymbiosis Microorganisms, including bacteria, rickettsia, mycoplasmas, viruses, and yeasts, live within the cells of many eukaryotic organisms including insects. Symbiosis often is used to mean mutualism, but originally included parasitism and mutualism. Intracellular symbionts have been called endocytobionts, with no assumptions being made about whether the relationship is mutualistic or parasitic (= endocytobiosis).

enhancer Sequences of DNA that can increase transcription of neighboring genes over long distances up or downstream of the gene and in either possible orientation.

enhancer trap A method to identify genes based on their pattern of expression. A reporter gene under the control of a weak constitutive promoter, when brought in proximity to a tissue-specific enhancer element, would be regulated by that enhancer, resulting in the expression of the reporter gene in a tissue- and stage-specific pattern similar to that of the native gene normally controlled by the enhancer.

environmental sex determination A method of sex determination in which the environment, such as temperature, has a significant effect on the developmental processes leading to one or the other sex.

enzyme A protein catalyst that is not itself used up in a reaction. Enzymes are produced by living cells to catalyze specific biochemical reactions. Enzymes may also contain nonprotein components called coenzymes that are essential for catalytic activity.

epistatic Epistasis is the nonreciprocal interaction of nonallelic genes. A gene epistatic to another masks the expression of the second gene.

Escherichia coli A bacterium that commonly inhabits the human intestine. Probably the most studied of all bacteria, it is used in many genetic experiments. Genetically engineered versions produce human proteins. (Its genome has been sequenced.)

EST Abbreviation for expressed sequence tags.

ethidium bromide A dye that binds to double-stranded DNA by intercalating between the stands. DNA stained with EtBr fluoresces under UV illumination.

euchromatin Regions of a eukaryotic chromosome that appear less condensed and stain less well with DNA-specific dyes than other segments of the chromosome.

eukaryote An organism with cells containing a membrane-bound nucleus that reproduces by meiosis. Cells divide by mitosis. Oxidative enzymes are packaged within mitochondria.

exogenous DNA DNA from an outside source. In genetic engineering, DNA from one organism is often inserted into another by a variety of methods.

exon One of the coding regions of a discontinuous gene.

exonuclease A nuclease, which degrades a nucleic acid molecule by progressive cleavage along its length, beginning at the $3'$ or $5'$ end.

expression vector Vectors that are designed to promote the expression of gene inserts. Usually an expression vector has the regulatory sequence of a gene ligated into a plasmid that contains the gene of interest. This gene lacks its own regulatory sequence. The plasmid with this new combination (regulatory sequence + gene) is placed into a host cell such as *E. coli* or yeast, where the protein product is produced.

extrachromosomal gene A gene not carried by the cell's chromosomes, such as mitochondrial or plasmid-borne genes.

F pili The presence of an F (fertility) factor determines the sex of a bacterium. Cells with F factors (circular DNA molecules that are about 2.5% of the length of the bacterial chromosome) are able to function as males, by producing an F pilus. The F pilus is a hollow tube through which chromosomal DNA is transferred during bacterial conjugation.

F statistics A set of coefficients that describe how genetic variation is partitioned within and among populations and individuals, such as F_{ST} and inbreeding coefficient.

F1 hybrid The first-generation offspring of a cross between two different strains.

facultatively heterochromatic Chromosomal material that, unlike euchromatin, shows maximal condensation in nuclei during interphase. Constitutive heterochromatin is composed of repetitive DNA, is late to replicate, and is transcriptionally inactive. Portions of the chromosome that are normally euchromatic may become heterochromatic at a particular developmental stage (= facultative heterochromatin). An example of facultative heterochromatin is the inactivated X chromosomes in the diploid somatic cells of mammalian females.

fate maps A technique used to analyze behavior in *Drosophila*. Using a ring X chromosome which is usually lost, individuals can be produced which are partly male and partly female. The pattern of genetic markers can be used to construct a fate map, which correlates precise anatomical sites on the embryonic blastoderm with abnormalities affecting behavior.

FB transposons A family of transposons in *Drosophila* that are associated with chromosomal abnormalities.

fertilization The union of the haploid male and female gametes to produce a diploid zygote, marking the start of the development of a new individual and the beginning of cell differentiation.

fibroin See silk.

fingerprinting DNA fingerprinting relies on the presence of simple tandem-repetitive sequences that are present throughout the genome. The regions show length polymorphisms, but share common sequences. DNA from different individuals is cut and separated by size on a gel. A probe containing the core sequence is used to label those fragments that contain the complementary DNA sequences. The pattern on each gel is specific for a given individual and can be used to establish parentage.

flanking sequence A segment of DNA that precedes or follows the region of interest on the molecule.

FLP recombinase Yeast FLP recombinase is able to catalyze recombination in which a DNA segment that is flanked by direct repeats of FLP target sites (FRTs) can be excised from the chromosome. If two homologous chromosomes each bear an FRT site, mitotic recombination can occur in *Drosophila*, leading to the introduction of DNA into known, and specific, sites. FRT sites can be introduced into *Drosophila* chromosomes by P-element-mediated transformation.

foldback DNA DNA that contains palindromic sequences that can form hairpin double-stranded structures when denatured DNA is allowed to renature.

forward genetics Analysis of the phenotype or function leads to identification of interesting mutants, which might be used to analyze a particular process or clone the genes responsible for regulating this process.

forward mutation A mutation from the wild type to the mutant. A **back mutation** restores the wild-type phenotype.

frameshift mutation A mutation resulting from inserting or deleting a group of nucleotides that is not a multiple of three, so that the polypeptide produced will probably have a new set of amino acids specified for downstream of the frameshift.

F_{ST} Coancestry coefficient; a measure of the relatedness of individuals.

functional genomics Study of what traits/functions are conferred on an organism by specific DNA sequences. Typically functional genomics occurs after the DNA sequences have been identified.

fusion protein A hybrid protein molecule produced when a gene of interest is inserted into a vector and displaces the stop codon for a gene already present in the vector. The fusion protein begins at the amino end with a portion of the vector protein sequence and ends with the protein of interest.

G-banding Dark bands on chromosomes produced by Giemsa staining; G-bands occur in A-T rich regions of the chromosome.

gamete A germ or reproductive cell, i.e., the sperm and ovum or egg.

gap genes Gap gene mutants lack large areas of the normal cuticular pattern. Three wild-type gap genes, *Krüppel*$^+$, *hunchback*$^+$, and *knirps*$^+$, regionalize the embryo by delimiting domains of homeotic gene expression and effect position-specific regulation of the pair-rule genes.

gating The process of shutting off a function when the value of a specific parameter attains a critical level.

gel electrophoresis Separation of molecules on the basis of their net electrical charge and size.

gene A segment of DNA that codes for an RNA and/or a polypeptide molecule. It includes regions preceding and following the coding region, as well as introns.

gene amplification The production of multiple copies of a DNA segment in order to increase the rate of expression of a gene carried by the segment. The chorion genes of *Drosophila* are amplified in the ovary.

gene boundaries Boundaries between active and inactive chromatin occur along the chromosomes. Such boundaries are established by **insulators** that act as a neutral barrier to the influence of neighboring elements.

gene cloning Insertion of a fragment of DNA containing a gene into a cloning vector and subsequent propagation of the recombinant DNA molecule in a host organism. Recently, cloning of a DNA fragment by the polymerase chain reaction has simplified the technology.

gene conversion A genetic process by which one sequence replaces another at an orthologous or paralogous locus, resulting in concerted evolution. May result from mismatch repair.

gene duplication The duplication of a DNA segment coding for a gene; gene duplication produces two identical copies which may retain their original function allowing the organism to produce larger amounts of a specific protein. Alternatively, one of the gene copies may be lost by mutation and become a pseudogene, or a duplicated gene can evolve to perform a different task.

gene expression The process by which the information carried by a gene is made available to the organism through transcription and translation.

gene gun A method for propelling microscopic particles coated with DNA into cells, tissues, and organelles to produce stable or transient transformation.

gene library A collection of recombinant clones derived from genomic DNA or from the cDNA transcript of an mRNA preparation. A complete genetic library is sufficiently large to have a high probability of containing every gene in the genome.

gene regulation The mechanisms that determine the level and timing of gene expression.

gene targeting A technique for inserting changes into a genetic locus in a desired manner. The desired locus is transferred into an embryo by microinjection where it is allowed to undergo homologous recombination into the chromosomes, replacing the original allele.

gene transfer The movement of a gene or group of genes from a donor to a recipient organism.

genetic code The rules that determine which triplet of nucleotides code for which amino acid during translation. There are more than 20 different amino acids and four bases (adenine, thymine, cytosine, and guanine). There are 64 potential combinations of the four bases in triplets ($4 \times 4 \times 4$). A doublet code would only be able to code for 16 (4×4) amino acids. Since only 20 amino acids exist, there is redundancy in the system so that some amino acids are coded for by two or three different triplets (codons).

genetic distance A measure of the evolutionary divergence of different populations of a species, as indicated by the number of allelic substitutions that have occurred per locus in the two populations. The most widely used measure of genetic distance is that of Nei (1972), $D = -\ln(I)$.

genetic diversity (**G_{ST}**) Variation in populations averaged over different loci.

genetic engineering The deliberate modification of genes by man. Also called gene splicing, gene manipulation, recombinant DNA technology.

genetic linkage Genes are located together on the same chromosome.

genetic marker An allele whose phenotype is recognized and which can be used to monitor the inheritance of its gene during genetic crosses between organisms with different alleles.

genetic sex determination system The mechanism in a species by which sex is determined. In most organisms, sex is genetically, rather than environmentally, determined.

genic balance model of sex determination in Hymenoptera Sex is determined by a balance between nonadditive male-determining genes and additive female-determining genes scattered throughout the genome. Maleness genes (m) have noncumulative effects but femaleness genes (f) are cumulative. In haploid individuals m > f, which results in a male, whereas in diploids ff > mm, which results in a female.

genome The total complement of DNA in an organism.

genomic footprinting A technique for identifying a segment of a DNA molecule in a living cell that is bound to some protein of interest. The phosphodiester bonds in the region covered by the protein are protected from attack by endonucleases. A control sample of pure DNA and one of protein-bound DNA are subjected to endonuclease attack. In DNA footprinting the resulting fragments are electrophoresed on a gel to separate them according to their size. For every bond that is susceptible to restriction, a band is found on the control gel. The gel prepared from the protein-bound DNA will lack bands, and these missing bands identify where the protein is protecting the DNA from being cut. The goal of genomic footprinting is to determine the contacts between DNA bases and specific proteins in a living cell. DNA footprinting determines these interactions *in vitro*.

genomic imprinting The process by which some genes are found to function differently when they are transmitted by the mother rather than the father, or vice versa. Mechanisms of imprinting may include methylation of the DNA. The more a gene is methylated, the less likely it is to be expressed.

genomic library A random collection of DNA fragments from a given species inserted into a vector (plasmids, phages, cosmids). The collection must be large enough to include all the unique nucleotide sequences of the genome.

genomics The study of genome data. The complete DNA sequences of organisms such as the human, mouse, rat, zebrafish, *D. melanogaster, C. elegans*, and *Arabidopsis thaliana* can provide a plethora of information on entire families of genes and whole pathways of interacting proteins. See also functional genomics, proteomics, and structural genomics.

genotype The genetic constitution of an organism. The phenotype of the organism is its appearance or observable character.

geotaxis The movement of an animal in response to gravity.

glycosylation A process in which a sugar or starch is linked to a protein molecule.

GMO Genetically modified organism.

guanine A purine in one of the nucleotides in DNA and RNA.

haploid Cells or organisms that contain a single copy of each chromosome.

Hardy–Weinberg equilibrium An equilibrium of genotypes achieved in populations of infinite size in which there is no migration, selection, or mutation after at least one generation of panmictic mating. With two alleles, *A* and *a,* of frequency *p* and *q*, the Hardy–Weinberg equilibrium frequencies of the genotypes *AA, Aa*, and *aa* are p^2, $2pq$, and q^2, respectively.

helicase The enzyme responsible for breaking the hydrogen bonds that hold the double helix together so that replication of DNA can occur.

helix A spiral staircase-like structure with a repeating pattern.

helper plasmid A plasmid that is able to supply something to a defective plasmid, thus enabling the defective plasmid to function.

heritability In the **broad sense** ($h_B = V_G/V_P$), the fraction of the total phenotypic variance that remains after exclusion of the variance due to environmental effects. In the **narrow sense**, the ratio of the additive genetic variance to the total phenotypic variance (V_A/V_P).

Hermes A transposable element that has been engineered for transforming insects other than *Drosophila*. *Hermes* was discovered in the house fly *Musca domestica*.

heterochromatin The regions of the chromosome that have large amounts of noncoding repetitive DNA.

heteroduplex DNA A hybrid DNA–DNA molecule formed from tracer and driver from different individuals or species.

heterogametic sex The sex that produces gametes containing unlike sex chromosomes. Many males are XY and thus heterogametic. Lepidopteran females are the heterogametic sex. Crossing over is often suppressed in the heterogametic sex.

heterogeneous nuclear ribonucleoproteins (hnRNPs) Pre-mRNAs and mRNAs are associated with a set of at least 20 proteins throughout their processing in the nucleus and transport to the cytoplasm. Some of these hnRNPs contain nuclear export signals.

heterologous DNA DNA from a species other than that being examined.

heterologous recombination Recombination between two DNA molecules that apparently lack regions of homology.

heteroplasmy The coexistence of more than one type of mitochondrial DNA within a cell or individual.

heterosis Also known as hybrid vigor.

heterozygosity Having a pair of dissimilar alleles at a locus (eg., Aa); a measure of genetic variation in a population estimated by a single locus or an average over several loci.

heterozygote A diploid cell or organism that contains two different alleles of a particular gene.

highly repetitive DNA DNA made up of short sequences, from a few to hundreds of nucleotides long, which are repeated on an average of 500,000 times.

histone gene family See histones.

histones Basic proteins that make up nucleosomes and have a fundamental role in chromosome structure.

Hogness box A DNA sequence 19–27 bp upstream from the start of a eukaryotic structural gene to which RNA polymerase II binds. The sequence is usually 7 bp long (TATAAAA); named in honor of D. Hogness. Often called TATA box and pronounced "tah-tah."

holocentric Chromosomes that have diffuse centromeres.

homeo domain See homeobox.

homeobox A conserved DNA sequence about 180 bp in size found in a number of homeotic genes involved in eukaryotic development. Homeobox genes (genes to which the homeobox is attached) are those genes that are responsible for embryonic development.

homeotic The replacement of one serial body part by a serially homologous body part.

homeotic gene Genes that determine the identification and sequence of segments during embryonic development in insects. Although most genes with a homeo domain are in the homeotic class, a few are found among the segmentation genes. Homeotic genes have been described in a variety of insects other than *Drosophila*, including *Musca, Aedes, Anopheles, Blatella,* and *Tribolium.*

homeotic mutations Mutations in which one developmental pattern is replaced by a different but homologous one. Homeotic mutations of *Drosophila* and other insects cause an organ to differentiate abnormally and form a homologous organ that is characteristic of an adjacent segment. Examples in *Drosophila* include *aristapedia* in which the antenna becomes leglike, and *bithorax* in which halteres are changed into winglike appendages.

homoduplex DNA molecules A double-stranded DNA molecule in which the two strands

come from different sources in DNA–DNA hybridization. Heteroduplex DNA will denature or melt into single strands at lower temperatures than homoduplex DNA from a single source.

homogametic sex The sex that produces gametes with only one kind of sex chromosome. The females of many insects are XX and thus homogametic.

homologous chromosomes Two or more identical chromosomes.

homologous genes Two genes from different organisms and therefore of different sequence that code for the same gene product.

homology Homology has been defined as "having a common evolutionary origin," but also is often used to mean "possessing similarity or being matched."

homoplasy Phenomena that lead to similarities in character states for reasons other than inheritance from a common ancestor, including convergence, parallelism, and reversal.

homozygous Diploid cells or organisms that contain two identical alleles of a particular gene.

horizontal gene transfer The transfer of genetic information from one species to another. Mechanisms and frequency are not well understood in insects.

hot-start PCR Hot start is a method to optimize the yield of desired PCR product and to suppress nonspecific amplification. This is done by withholding an essential component of the PCR, such as the DNA polymerase, until the reaction mixture has been heated to a temperature that inhibits nonspecific priming and primer extension. See also polymerase chain reaction (PCR).

housekeeping genes Genes whose products are required by the cell for normal maintenance.

humoral immunity The immune system response that consists of soluble blood serum components that fight an infection.

hybrid dysgenesis A syndrome of genetic abnormalities that occurs when hybrids are formed between strains of *Drosophila melanogaster*, one carrying (P) and the other lacking (M) the transposable *P* element. The abnormalities include chromosomal damage, lethal and visible mutations, and sometimes sterility. Dysgenesis is caused by crossing P males × M females, but the reciprocal cross is not dysgenic.

hybridization probe A labeled nucleic acid molecule used to identify complementary or homologous molecules through the formation of stable base pairs.

hydrogen bonding A hydrogen bond is a weak electrostatic attraction between an electronegative atom (such as oxygen or nitrogen) and a hydrogen atom attached to a second electronegative atom. In effect, the hydrogen atom is shared between the two electronegative atoms.

hypertranscription Transcription of DNA at a rate higher than normal. For many species with an XY sex-determination system, the male compensates for his single X chromosome by hypertranscribing the X chromosome. He produces a nearly equal amount of gene product compared to what is produced by females with two X chromosomes.

imaginal discs Cells set off during embryonic development that will give rise, during the pupal stage, to adult organs.

***in silico* biology** *In silico* biology refers to the use of computers to perform biological studies.

***in situ* hybridization** The pairing of complementary DNA and RNA strands, or the pairing of complementary DNA single strands to produce a hybrid molecule in intact chromosomes or cells. Pairing is detected by some form of label. For chromosomal squash preparations on glass slides, the DNAs are denatured and adhering RNAs and proteins are removed. Then the DNA is incubated with tritium-labeled nucleic acid probes. The radioactive molecules hybridize with the DNA segments of specific chromosomal areas and these are visualized in autoradiographs. It also can be used to identify DNA sequences in DNAs released from lysed bacterial colonies onto nitrocellulose filters.

***in vitro* packaging** The production of infectious particles by enclosing naked DNA in

lambda (λ) phage packaging proteins and preheads.

inbreeding coefficient The correlation of genes within individuals (F_{IT}), or the correlation of genes within individuals within populations (F_{IS}). Both F_{IS} and F_{IT} are measures for deviation from expected Hardy–Weinberg proportions.

indel An insertion or deletion in a DNA sequence.

independent assortment See Law of Independent Assortment.

inducible enzymes Enzymes whose rate of production is increased by the presence of certain molecules.

initiation codon AUG serves as an initiation codon when it occurs at the start of a gene; it marks the site where translation should begin. AUG also codes for methionine, so most newly synthesized polypeptides will have this amino acid at the amino terminus, although it may later be removed by posttranslational processing of the protein. AUG is the only codon for methionine, so AUGs that are not initiation codons are also found in the middle of a gene.

insertion mutation Alteration of a DNA sequence by inserting one or more nucleotides.

insertion sequences Insertion sequences are the simplest transposable elements, carrying no genetic information except what is needed to transpose (i.e., transposase). Usually 700–2500 bp long, denoted by the prefix IS and followed by the type number.

insertion vectors Vectors that have a single target site at which foreign DNA is inserted.

insulators Novel sequence elements found recently in *Drosophila* that are associated with boundaries between active and inactive genes, protecting against position. Insulators act as a neutral barrier against both positive and negative effects of the chromosomal environment.

intercalating agent A chemical compound which is able to invade the space between adjacent base pairs of a double-stranded DNA molecule; including ethidium bromide.

intergenic region The noncoding region between segments of DNA that code for genes.

interphase The stage of the cell cycle when chromosomes are not visible by light microscopy. During interphase, DNA synthesis occurs.

introgression The incorporation of genes of one species into the gene pool of another. If the ranges of two species overlap and fertile hybrids are produced, they will tend to backcross with the more abundant species.

intron A region of eukaryotic DNA coding for RNA that is later removed during splicing; it does not contribute to the final RNA product.

inverse PCR Inverse PCR allows amplification of an unknown DNA sequence that flanks a "core" region with a known sequence. The basic method for inverse PCR involves digesting template DNA, circularizing the digested DNA, and amplifying the flanking DNA outside the core region with the primers oriented in the opposite direction of the usual orientation. Primers for inverse PCR are synthesized in the opposite orientation and are homologous to the ends of the core region so that DNA synthesis proceeds across the *uncharacterized* region of the circle rather than across the characterized core region.

inversion Alteration of the sequence of a DNA molecule by removal of a segment followed by its reinsertion in the opposite orientation.

inverted repeat Two identical nucleotide sequences repeated in opposite orientation in a DNA molecule, either adjacent to one another or some distance apart.

ion channels The membrane passages that allow certain ions to cross the membrane.

ionic selectivity The ability of ion channels to permit certain ions to cross the membrane, but not others.

isozymes (isoenzymes) Multiple forms of an enzyme that differ from each other in their substrate affinity, in their activity, or in their regulatory properties. Isozymes are complex proteins of paired polypeptide subunits. They often have different isoelectric points and can be separated by electrophoresis.

jumping genes Genes that move within the genome, usually because they are associated with transposable elements.

junk DNA The proportion of DNA in a genome that *apparently* has no function. Also called parasitic or selfish DNA.

kilobase A kilobase (kb) of DNA = 1000 nucleotides.

kilodalton (kDa) A unit of mass equal to 1000 daltons (Da). One dalton is nearly equal to the mass of a hydrogen atom.

kin selection A theory put forth by W. D. Hamilton (1964) that states that an altruistic act is favored because it increases the inclusive fitness of the individual performing the social act. Inclusive fitness is the fitness of the individual as well as his effects on the fitness of any genetically related neighbors. Kin selection could explain the evolution of sociality, which appears to have developed as many as 11 times in the order Hymenoptera.

Klenow fragment A portion of bacterial DNA polymerase I derived by proteolytic cleavage. It lacks the 5′-to-3′ exonuclease activity of the intact enzyme.

lagging strand The DNA strand in the double helix which is copied in a discontinuous manner during DNA replication; short segments of DNA produced during the replication are called Okazaki fragments.

lambda or λ A double-stranded DNA virus (bacteriophage) that can invade *E. coli*. Once inside the cell λ can enter a lysogenic cycle or a lytic cycle of replication, which results in death of the host cell. λ has been genetically engineered as a vector for cloning. λ is also a microliter unit of measurement, the volume contained in a cube 1 mm on a side.

Law of Independent Assortment One of Mendel's laws. The members of different pairs of factors assort independently. Different pairs of alleles assort independently into gametes during gametogenesis, if they are on different chromosomes. The subsequent pairing of male and female gametes is at random, which results in new combinations of alleles.

Law of Segregation One of Mendel's laws. The factors of a pair of characters segregate. Separation into different gametes, and thus into different progeny, of the two members of each pair of alleles possessed by the diploid parent.

leader sequence An untranslated segment of mRNA from its 5′ end to the start codon.

leading strand The DNA strand in the double helix which is copied in a continuous fashion during DNA replication.

lethal mutation Mutation of a gene to yield no product, or a defective gene product, resulting in the death of the organism because the gene product is essential to life.

leucine zipper DNA binding proteins that contain four to five leucine residues separated from each other by six amino acids. The leucines on two protein molecules interdigitate and dimerize in a specific interaction with a DNA recognition sequence. Leucine zippers are involved in regulating gene expression.

library A set of cloned DNA fragments which represent the entire genome.

ligase DNA ligases are enzymes that catalyze the formation of a phosphodiester bond between adjacent 3′-OH and 5′-P termini in single-stranded DNA. DNA ligases function in DNA repair to seal single-stranded nicks between adjacent nucleotides in a double-stranded DNA molecule.

ligation Enzymatic joining together of nucleic acid molecules through their ends.

likelihood methods Likelihood methods of analyzing DNA sequence data rely on genetic models and provide a basis for statistical inference. Maximum likelihood methods of tree construction assume the form of the tree and then choose the branch length to maximize the likelihood of the data given that tree. These likelihoods are then compared over different possible trees and the tree with the greatest likelihood is considered to be the best estimate.

linkage A linkage group is a group of genes located on a single chromosome.

linkage map A diagram of the order and relative distances between gene loci on chromosomes, based on the frequency of recombination of the linked genes in the genomes of progeny obtained from crossing parents with different genetic markers.

linker DNA The DNA that links nucleosomes; the function of linker DNA is unresolved.

locus The position of a gene on a chromosome. Plural: loci.

long germ band development A pattern of development in insects such as *D. melanogaster* in which the pattern of segmentation is established by the end of blastoderm.

long-period interspersion genome organization This organization of the DNA in the genome involves long (>5600 bp) repeats alternating with very long (>12 kb) uninterrupted stretches of unique DNA sequences. Long-period interspersion is characteristic of species with small genomes. Short-period interspersion involves a pattern of single-copy DNA, 1000–2000 bp long, alternating with short (200–600 bp) and moderately long (1000–4000 bp) repetitive sequences, which is characteristic of the DNA in most animal species.

lysis The process of disintegrating a cell, which involves rupturing the membranes, breaking up the cell wall and nuclear membrane.

lysogenic During the lysogenic phase of a bacteriophage, the DNA of a virus is integrated into the chromosome of its bacterial host.

lytic A virus in a lytic phase undergoes intracellular multiplication, and lysis of the bacterial host cell results.

major groove The larger of the two grooves that spiral around the surface of the double helix of the DNA molecule.

map unit In linkage maps, a 1% recombination frequency is defined as a map unit or one centimorgan. A number proportional to the frequency of recombination between two genes.

mariner A transposable element that has been engineered for transforming insects other than *Drosophila*. *mariner* elements are widely found in arthropods and in insect-parasitic nematodes, other nematodes, flatworms, hydras, humans, mouse, rat, Chinese hamster, sheep, and cows. *mariner* has been used to transform chicken, zebrafish, and a protozoan.

marker (DNA size marker) A DNA fragment of known size used to calibrate an electrophoretic gel.

marker (genetic) A trait that can be observed to occur (or not) in an organism. Marker genes include genes conferring resistance to antibiotics, expression of green fluorescent protein, eye color, etc.

maternal effect gene Genes with a maternal effect are genes in the mother which have an effect on the phenotype of her progeny. Usually the result of depositing products or maternally derived mRNAs in the egg that are used or transcribed by the embryo.

maternally inherited Characters that are transmitted primarily by cytoplasmic genetic factors (including mitochondria, viruses, some mRNAs) derived solely from the maternal parent. Also known as cytoplasmic inheritance or extranuclear heredity.

Maxam and Gilbert sequencing method A "chemical" method to sequence DNA developed in 1977 by A. M. Maxam and W. Gilbert. Single-stranded DNA derived from double-stranded DNA and labeled at the 5′ end with ^{32}P is subjected to several chemical cleavage protocols to selectively make breaks on one side of a particular base. The fragments are separated by size by electrophoresis on acrylamide gels and identified by autoradiography.

maximum parsimony methods Taxonomic methods that focus on the character values observed and minimizing the number of changes in character state between species over the tree, making the assumption that there have been approximately constant rates of change. The changes at each node in the tree are inferred to be those that require the least number of changes to give each of the two character states of the immediate descendants.

median melting temperature The temperature at which 50% of the double helices have denatured; the midpoint of the temperature range over which DNA is denatured.

meiosis The sequence of events occurring during two cell divisions to convert diploid cells into haploid cells.

meiotic drive Any mechanism that results in the unequal recovery of the two types of gametes produced by a heterozygote.

melting of DNA Melting DNA means to denature it by heat, breaking the hydrogen bonds that hold the two strands together.

messenger RNA (mRNA) RNA molecules which code for proteins and which are translated on the ribosomes.

methylation In bacteria, enzymes (modification methylases) that bind to the DNA attach methyl groups to specific bases. This methylation pattern is unique to and protects the species from its own restriction endonucleases. Methylation also occurs in eukaryotes and may be involved in genomic imprinting. Genes that are methylated are less likely to be active.

M13 bacteriophage A single-stranded bacteriophage cloning vehicle, with a closed circular DNA genome of approximately 6.5 kb. M13 produces particles that contain ss DNA that is homologous to only one of the two complementary strands of the cloned DNA and therefore is particularly useful as a template for DNA sequencing.

M13 universal primer A primer derived from the M13 bacteriophage is used for sequencing reactions and has been used to identify satellite DNA sequences in many organisms.

μg A microgram (μg) is 10^{-6} of a gram.

micron One-thousandth of a millimeter. A Greek letter μ is its symbol.

microsatellite DNA Pieces of the same small segment which are repeated many times.

minor groove The smaller of the two grooves that spiral around the surface of the DNA double helix.

Minos A transposable element that has been engineered for transforming insects other than *Drosophila*. *Minos* has a wide host range and can transform human cell lines, making it potentially useful for mutagenesis and analysis of the human genome.

mitochondrion An organelle that occurs in the cytoplasm of all eukaryotes, capable of self-replicating. Each mitochondrion is surrounded by a double membrane. The inner membrane is highly invaginated, with projections called cristae that are tubular or lamellar. Mitochondria are the sites of oxidative phosphorylation which result in the formation of ATP. Mitochondria contain distinctive ribosomes, transfer RNAs, and aminoacyl-tRNA synthetases. Mitochondria depend upon genes within the nucleus of the cells they inhabit for many essential mRNAs. Proteins translated from mRNAs in the cytoplasm are imported into the mitochondrion. Mitochondria are thought to be endosymbionts derived from aerobic bacteria that are associated with primitive eukaryotes. The genetic code of mitochondria differs slightly from the universal genetic code. Mitochondria are transferred primarily through the egg, and thus are maternally inherited.

mitosis The sequence of events that occur during the division of a single cell into two daughter cells.

mobile genetic element See transposable element.

moderately repetitive DNA Nucleotide sequences that occur repeatedly in chromosomal DNA. Repetitive DNA is moderately (= middle) repetitive or highly repetitive. Highly repetitive DNA contains sequences of several nucleotides repeated millions of times. It is a component of constitutive heterochromatin. Middle-repetitive DNA consists of segments 100–500 bp long repeated 100 to 10,000 times each. This class also includes the genes transcribed into tRNAs and rRNAs.

molecular biology A term broadly used to describe biology devoted to the molecular nature of the gene and its biochemical reactions such as transcription and translation.

molecular clock The hypothesis that molecules evolve in direct proportion to time so that differences between molecules in two different species can be used to estimate the time elapsed since the two species last shared a common ancestor.

molecular evolution That subdivision of the study of evolution that studies the structure and functioning of DNA at the molecular level over time.

molecular genetics Genetic studies that focus on the molecular nature of genes and gene expression.

molecular phylogeny An analysis of the relationships of groups of organisms as reflected by

the evolutionary history detected in molecules (proteins, DNA).

molecular systematics The detection, description, and explanation of molecular diversity within and among species.

morphogen Molecules whose local concentration directly determines the local pattern of differentiation.

mRNA Messenger RNA.

mtDNA Mitochondrial DNA.

monoclonal antibody A single antibody produced in quantity by cultured hybridoma cell lines.

Muller's ratchet The accumulation of deleterious mutations that can lead to extinction of a population of a sexual species.

multigene family A group of genes that are related either in nucleotide sequence or in terms of function; they are often clustered together.

multiple-locus, multiple-allele model A model for sex determination in Hymenoptera.

multiplex PCR When more than one pair of primers is used in a PCR, multiple segments of target DNA can be amplified simultaneously and thus conserve template, save time, and minimize expense. See also polymerase chain reaction (PCR).

mushroom body Two nerve structures in the brain of insects.

mutagen A chemical or physical agent able to induce a mutation in a DNA molecule.

mutant An organism expressing the effects of a mutated gene in its phenotype.

mutation A change in the nucleotide sequence of a DNA molecule. Mutations can involve duplications, deletions, inversions, translocations, and substitutions.

nanogram (ng) A nanogram is one billionth of a gram.

nanometer (nm) A nanometer is one billionth of a meter.

negative heterosis The inferiority of a heterozygote over that of the homozygotes with respect to one or more traits such as growth, survival, or fertility.

neuropeptides Small molecules functioning within and without the nervous system of insects to modify behavior.

neutral theory of molecular evolution A theory that the majority of the nucleotide substitutions in the course of evolution are the result of the random fixation of neutral or nearly neutral mutations, rather than the result of positive Darwinian selection. Many protein mutations are selectively neutral and are maintained in the population by the balance between new mutations and their random extinction. Neutral mutations have a function, but they are equally effective in comparison to the ancestral alleles in the survival and reproduction of the organisms carrying them. Neutral mutations spread within populations by chance because only a relatively small number of gametes are sampled each generation and thus are transmitted to the next generation.

nick A break in a single strand of a double-stranded DNA molecule.

nick translation A commonly used method of labeling DNA molecules with radioactive isotopes. DNA polymerase I is used to incorporate radiolabeled nucleotides in an *in vitro* reaction.

nitrogenous base A purine or pyrimidine compound that forms part of the structure of a nucleotide.

noncoding strand The polynucleotide of the DNA double helix that does not carry the genetic information, but that is the complement of the coding strand.

nonsense mutation A mutation in a nucleotide sequence that changes a triplet coding for an amino acid into a termination codon so that a truncated polypeptide is produced which can alter the protein's activity.

Northern blotting A technique for transferring mRNAs from an agarose gel to a nitrocellulose filter paper sheet via capillary action. The RNA segment of interest is probed with a radiolabeled DNA fragment or gene.

nuclear genome The portion of the genome contained in the nucleus of eukaryotes on chromosomes.

nuclear pore complex A large structure forming a transport channel through the nuclear envelope.

nucleic acid Either of the polymeric molecules DNA or RNA.

nucleic acid hybridization The bonding of two complementary DNA strands, or one DNA and one RNA strand, to identify nucleic sequences of interest. Southern blot, Northern blot, and plaque or colony hybridization techniques are all based on nucleic acid hybridization. All employ labeled probes to identify DNA or RNA of interest.

nucleolus A nucleolus is an RNA-rich, spherical body associated with a specific chromosomal segment, the nucleolus organizer. The nucleolus organizer contains the ribosomal RNA genes and the nucleolus is composed of the primary products of these genes, their associated proteins, and a variety of enzymes.

nucleoside A compound consisting of a purine or pyrimidine base attached to a five-carbon sugar.

nucleosome A basic structure by which eukaryotic chromosomes are organized and compacted. Nucleosomes comprise an octamer of histone proteins with DNA coiled around them and are connected to other nucleosomes by linker DNA.

nucleotide A compound consisting of a purine or pyrimidine base attached to a five-carbon sugar, to which a mono-, di-, or triphosphate is attached. A monomeric unit of DNA or RNA.

nucleus The membrane-bound structure of a eukaryotic cell containing the DNA organized into chromosomes.

null allele An allele that produces no functional product and therefore usually behaves as a recessive.

odorant binding protein A protein that enhances the ability to smell odorants in small quantities—quantities lower than those needed to activate olfactory nerves.

Okazaki fragments Short fragments of DNA that are synthesized during replication of the lagging strand of the DNA molecule.

oligo See oligonucleotide.

oligonucleotide Short chains of single-stranded DNA or RNA nucleotides that have been synthesized by linking together a number of specific nucleotides. Used as synthetic genes or DNA probes.

oocytes Cells produced by the ovaries that eventually become an ovum (egg cell) after meiosis.

open reading frame (ORF) A series of codons with an initiation codon at the 5′ end. Often considered synonymous with "gene" but used to describe a DNA sequence that looks like a gene but to which no function has been assigned.

origin of replication (ORI) A base sequence in DNA that is recognized as the position at which the replication of DNA should begin. In eukaryotes, multiple origins of replication occur on each chromosome.

***P* element** *P* elements are transposable DNA elements first found in *Drosophila melanogaster*, where they can cause hybrid dysgenesis if P-containing strains are crossed with M strains lacking *P* elements. *P* elements have been engineered to serve as vectors to insert DNA into the germ line of *Drosophila* embryos.

pair-rule genes Mutated pair-rule genes result in repetitive aberrations throughout the germ band, with the removal of integral, alternate segment-width areas. The pair-rule genes (including *runt*$^+$, *hairy*$^+$, *fushi tarazu*$^+$, *even skipped*$^+$, *paired*$^+$, *odd-paired*$^+$, *odd-skipped*$^+$, *sloppy-paired*$^+$) are transiently expressed in seven or eight stripes during cellularization of the blastoderm.

palindrome A DNA sequence which reads the same in both directions taking account of the two strands, i.e., 5′-AAAAATTTTTT-3′
3′-TTTTTTAAAAAA-5′

paralogy Homology that arises via gene duplication.

parasegment The visible cuticular patterns of sclerites and sutures in an insect do not represent the embryonically determined true segments. Rather, the visible "segments" are parasegments.

parental imprinting (also genomic imprinting) The degree to which a gene expresses itself depends on which parent transmits the trait

to the progeny. Imprinting may result from different patterns of DNA methylation which occur during gametogenesis in the two sexes. For such a system to maintain itself generation after generation, it would have to be reversible.

parsimony Parsimony dictates that the minimal number of assumptions are made in an analysis.

PAS domain Protein sequence associated with signaling pathways that transmit environmental information (such as oxygen and light). Sometimes associated with protein–protein interactions.

paternal sex ratio (PSR) The PSR condition is only carried by males of the parasitic wasp *Nasonia vitripennis* and is transmitted via sperm to fertilized eggs. After an egg is fertilized by a PSR-bearing sperm, the paternally derived chromosomes condense into a chromatin mass and subsequently are lost. The PSR chromosome itself survives, disrupting normal sex determination by changing fertilized diploid (female) eggs into haploid PSR males. PSR is the first known B chromosome of its kind and is unusual in its ability to destroy the genome of its carrier each generation

pathogen A virus, bacterium, parasitic protozoan, or other microorganism that causes disease by invading the body of a host; infection is not always disease because infection does not always lead to injury of the host.

PCR See polymerase chain reaction (PCR).

PCR-RFLP A technique that combines the PCR and RFLP analysis. Genomic DNA is amplified by traditional PCR. Once the DNA is amplified, it is cut with restriction enzymes, electrophoresed, and visualized by ethidium bromide staining. Because the DNA was amplified by the PCR, the DNA fragments can be visualized without having to blot and probe with a labeled probe, thus making PCR-RFLP more sensitive and inexpensive than traditional RFLP analysis.

peptide bond The chemical bond that links adjacent amino acids into a polypeptide.

phage (bacteriophage) A virus that attacks bacteria. Frequently used as vectors for carrying foreign DNA into cells by genetic engineers.

phagemid A phagemid is a hybrid vector molecule engineered from plasmid and M13 vectors. Phagemids provide a method for obtaining single-stranded DNA because they contain two replication origins, one a standard plasmid origin that allows production of ds DNA, and the other from M13, which allows the synthesis of ss DNA if the host cell is superinfected with a helper phage.

phenetic systematics Classification based on overall similarities among living organisms. All possible characters are examined and average similarities are calculated, with all characters assumed to be of equal importance.

phenogram A branching diagram that links different taxa by estimating overall similarity based on data from characters. Characters are not evaluated as to whether they are primitive or derived.

phenomics The study of phenotypes with knowledge of the genotypes.

phenotype The observable characteristics of an organism that are determined by both genotype and environment.

pheromone-binding protein Two soluble proteins are found in the lymph, a pheromone-degrading esterase and a pheromone-binding protein. The pheromone-binding proteins bind species-specific pheromones and are present in very high concentrations. Volatile hydrophobic odorant molecules have to enter an aqueous compartment and traverse a hydrophilic barrier before reaching olfactory neurons. The function of the pheromone-binding proteins is not fully resolved, although they are thought to be involved in carrying the hydrophobic odorant through the sensillum lymph toward the receptor proteins located in the dendrite membranes.

phosphodiester bond The chemical bond that links adjacent nucleotides in a polynucleotide.

phosphorylation The combination of phosphoric acid with a compound. Many proteins in eukaryotes are phosphorylated.

phototaxis The movement of a cell or organisms toward or away from light.

phyletic speciation The gradual transformation of one species into another without an increase in species number at any time within the lineage. Also called vertical evolution or speciation.

phylogenetic tree A graphic representation of the evolutionary history of a group of taxa or genes.

phylogenetics The reconstruction of the evolutionary history of a group of organisms or genes.

phylogeny The evolutionary history of a group of taxa or genes, and their ancestors.

physical map A map of the order of genes on a chromosome. The locations are determined by DNA sequencing, producing overlapping deletions in polytene chromosomes, or electron micrographs of heteroduplex DNAs.

picogram A picogram is 10^{-12} gram. A picogram of DNA is approximately 0.98×10^9 base pairs.

piggyBac A transposable element that has been engineered for transforming insects other than *Drosophila*.

plaque A clear spot on an opaque bacterial lawn in a petri dish. A plaque results after a single phage adsorbs to a bacterial cell, infects it, and lyses, releasing progeny phage. The progeny phage infect nearby bacteria and produce more phage until a clear area becomes visible to the naked eye. Each clear area contains many copies of a single phage and, if the phage is a vector containing exogenous DNA, it contains many copies of the foreign DNA.

plaque hybridization See plaque screening.

plaque screening Plaque screening is employed to identify, by nucleic acid hybridization with radiolabeled probes, those plaques containing specific DNA sequences.

plasmid Circular, ds DNA molecules found in bacteria that are often used in cloning. Plasmids are independent, stable, self-replicating, and often confer resistance to antibiotics. Often used in recombinant DNA work as vectors of foreign DNA.

pleiotropic Term used to describe a gene that affects more than one, apparently unrelated, trait.

plesiomorphic A character used to reconstruct a phylogeny that is ancestral or primitive.

point mutation A mutation that results from changes in a single base pair in a DNA molecule.

pole cells The precursors of the germ cells become separated early in embryonic development in *D. melanogaster* into distinctive cells in the posterior of the egg.

poly-A tail The processing of the $3'$ end of the pre-mRNA molecule by the addition of as many as 200 adenine nucleotides, which may determine mRNA stability.

polyacrylamide gel Polyacrylamide gels result from the polymerization of acrylamide monomers into linear chains and the linking of these chains with *N*,*N*$'$-methylenebisacrylamide (bis). The concentration of acrylamide and the ratio of acrylamide to bis determine the pore size of the three-dimensional network and its sieving effect on nucleic acids of different size.

polyacrylamide gel electrophoresis (PAGE) Process by which molecules are separated based on size and charge using a polyacrylamide gel and electrical current.

polydnaviruses The polydnaviruses are viruses with double-stranded, circular DNA genomes and found only within certain groups of parasitic Hymenoptera. Virus particles replicate only in the wasp ovary and are secreted into the oviducts. During oviposition, virus is injected into host larvae. It is believed that one (or more) gene in the virus contributes to the immunosuppressive state of the host, thus allowing the parasitoid eggs and larvae to survive. The polydnaviruses appear to integrate into parasitoid chromosomal DNA, but are also present in extrachromosomal molecules.

polylinker A genetically engineered segment in a vector that allows exogenous DNA to be cloned into that region by one of two or more unique restriction sites.

polymer A chemical compound constructed from a long chain of identical or similar units.

polymerase chain reaction (PCR) A method for amplifying DNA by means of DNA polymerases such as *Taq* DNA polymerase. PCR fundamentally involves denaturing double-stranded DNA, adding dNTPs, DNA polymerase, and primers. DNA synthesis occurs, resulting in a doubling of the number of DNA molecules defined by the primers. Additional rounds of denaturation and synthesis occur, resulting in a geometric increase in DNA molecules because each newly synthesized molecule can serve as the template for subsequent DNA amplification. Modifications of PCR primers have been developed for special purposes. The PCR is used to clone genes, produce probes, produce ss DNA for sequencing, and carry out site-directed mutagenesis. DNA sequence differences are used to identify individuals, populations, and species.

polymorphism Two or more genetically different classes in the same interbreeding population.

polynucleotide A polymer consisting of nucleotide units.

polypeptide (protein) A chain of amino acids linked by peptide bonds; each protein is a gene product.

polyploidy An increase in the number of copies of the haploid genome. Most individuals are 2n, but species are known that are polyploid (3n, 4n, 5n, 6n) and such species are parthenogenetic because of the difficulty of maintaining normal meiosis. Many insect species have tissues that are polyploid, including the salivary glands, nurse cells of the ovary, and fat body, but the germ-line tissues remain 2n.

polyribosome (polysome) An mRNA molecule in the process of being translated by multiple ribosomes.

polytene chromosomes Chromosomes in which the chromatid has duplicated up to 1000-fold without separating. Salivary gland chromosomes in *Drosophila* and other Diptera are polytene. The discrete bands of polytene chromosomes allow a physical map of genes to be constructed using light microscopy.

polyteny See polytene chromosomes.

position effect variegation The change in the expression of a gene when it is moved to a different region of the genome. The change in expression can be stable or variegated. Variegated position effects usually involve the suppression of wild-type gene activity when it is placed in contact with heterochromatin because of a chromosomal mutation. Under some conditions the gene may escape suppression and the final phenotype of the organism may be variegated, with patches of normal and mutant tissues.

positive and negative selection A method for detecting and obtaining, from among many cells or organisms, those few with the desired genetic changes induced by genetic engineering. Marker genes are inserted into the organism along with the desired genes; such marker genes confer resistance to antibiotics or other chemicals and allow researchers to identify those cells/individuals that contain the newly inserted genes.

posttranscriptional processing Changes made to mRNAs, rRNAs, and tRNAs before they are finished products.

posttranslational processing Changes to polypeptide chains after they have been synthesized—cleavage of specific regions to convert proenzymes to enzymes, phosphorylation, etc.

postzygotic isolating factors Factors that help to maintain reproductive isolation between species even if mating between them does occur, such as hybrid inviability or hybrid sterility.

pre-mRNA The unprocessed transcript of a protein-coding gene.

prezygotic isolating factors Aspects of a species' biology that help to maintain reproductive isolation so that mating between different species/populations does not occur, including mating discrimination or differences in habitat preferences.

primary transcript The immediate product of transcription of a gene or group of genes which will be processed to give the mature transcript(s).

primase The RNA polymerase that synthesizes the primer needed to initiate replication of a DNA polynucleotide during DNA replication.

primer A short oligonucleotide that is attached to a ss DNA molecule in order to provide a site at which DNA replication can begin.

primer-dimer artifacts Low molecular weight DNA products produced during PCR as artifacts when the reaction is carried out with high primer concentrations, too much DNA polymerase in early cycles, and small amounts of template DNA. The primer-dimer is made when the DNA polymerase makes a product by reading from the 3′ end of one primer across to the 5′ end of the other primer. This results in a sequence being produced that is complementary to each primer and can serve as a template for additional primer binding and extension.

prion Proteinaceous molecules found in the membranes of cells in the brains of vertebrates. In 1982, Stanley Prusiner discovered that mutated versions could cause a neurodegenerative disease called bovine spongiform encephalopathy (BSE or "mad cow disease") in cattle and Creutzfeld–Jakob Disease and kuru in humans. These "proteinaceous infected particles" do not contain DNA but are able to transmit the disease.

probe A probe is a molecule labeled with radioactive isotopes or another tag that is used to identify or isolate a gene, gene product, or protein.

prokaryote An organism whose cells lack a distinct nucleus.

promoter A region of DNA crucial to the accuracy and rate of transcription initiation. Usually immediately upstream of the gene itself.

proofreading A mechanism by which errors in DNA synthesis are corrected. Proofreading is carried out by a 3′ to 5′ exonuclease and increases the fidelity of the base-pairing mechanism.

protease An enzyme that degrades proteins.

protein The polymeric compounds made up of amino acids.

proteoglycan A protein that is glycosylated to a variety of polysaccharide chains.

proteome The protein complement of a cell.

proteomics The science and process of analyzing all the proteins encoded by a genome (a proteome). Currently the majority of all known and predicted proteins have no known cellular function. Determining protein function on a genome-wide scale can provide critical clues to the metabolism of cells and organisms. Proteomics involves understanding the biochemistry of proteins, processes, and pathways. Two-dimensional gel analyses were used in the late 1970s to identify proteins active (expressed) in different tissues at different times. Now, biological mass spectrometry is a powerful method for protein analysis, involving identification or localization of proteins and interactions of proteins.

proteosome A large protein complex in the cytoplasm of eukaryotic cells that contains proteolytic enzymes. Proteosomes break down proteins that have been tagged for destruction by the addition of ubiquitin.

pseudogene A nucleotide sequence that is similar to a functional gene, but without accurate information so that it is not functional.

PSR See paternal sex ratio (PSR).

puffing A swelling in the giant polytene chromosomes of salivary glands of many dipterans.

pulsed field gel electrophoresis A technique for separating DNA molecules by subjecting them to alternately pulsed, perpendicularly oriented electrical fields. The technique allows separation of the yeast genome into a series of intact chromosomes on a gel. Chromosomes larger than yeast chromosomes are digested with a restriction enzyme before electrophoresis.

purine One of the two types of nitrogenous bases that are components of nucleotides.

pyrimidine One of the two types of nitrogenous bases that are components of nucleotides.

Q-banding Bands on chromosomes produced by quinacrine staining. The staining can only be seen under UV light and is brightest in AT-rich regions.

quantitative genetics Analysis of the genetic influence of many genes and substantial environmental variation. It is assumed that Mendel's laws of discrete inheritance apply to complex characteristics, so that many genes, each with small effect, combine to produce observable differences among individuals in a

population. Quantitative genetics determines the sum of heritable genetic influence on traits, regardless of the complexity of genetic modes of action or the number of genes involved. It does not tell us which genes are responsible for the trait.

quantitative trait loci (QTL) Specific DNA sequences that are related to (located near to) known traits, which may be determined by multiple loci.

radiolabeling The attachment of a radioactive atom to a molecule; incorporation of ^{32}P-dNTPs into DNA.

RAPD-PCR RAPD is derived from the term Random Amplified Polymorphic DNA. PCR using single primers of arbitrary nucleotide sequence consisting of 9 or 10 nucleotides with a 50 to 80% G+C content, and no palindromic sequences. These 10-mers can act as a primer in PCR and yield reproducible polymorphisms from random segments of genomic DNA.

reading frame A nucleotide sequence from which translation occurs.

real-time PCR Real-time PCR is used to quantify gene expression using a fluorescence-detecting thermocycler to amplify specific sequences and measure their concentration simultaneously. See also polymerase chain reaction (PCR).

recessive A trait or gene is recessive if it is expressed in homozygous, but not heterozygous, condition.

reciprocal cross Crosses between individuals from two different strains (A, B), e.g., A × B and B × A.

recombinant DNA molecule A DNA molecule created by combining DNA fragments that are not normally contiguous.

recombinant DNA technology All the techniques involved in the construction, study, and use of recombinant DNA molecules. Often abbreviated rDNA, which can be confused with ribosomal DNA (rDNA).

recombination A physical process that can lead to the exchange of segments of two DNA molecules and that can result in progeny from a cross between two different parents with combinations of alleles not displayed by either parent.

redundancy Some amino acids have more than one codon. There are 64 possible combinations of four bases arranged in a triplet codon, but only about 20 amino acids.

regulatory gene A gene that codes for a protein that is involved in the regulation of the expression of other genes.

regulatory mutation Mutations that affect the ability to control gene expression.

regulatory sequence A DNA sequence involved in regulating the expression of a gene (a promoter or operator).

reinforcement An event (reward or punishment) that follows a response and increases or decreases the likelihood that it will recur.

repetitive DNA DNA sequences that are repeated a number of times in a DNA molecule or in a genome. Some repetitive DNA is associated with heterochromatin, centromeres, and telomeres. Middle-repetitive DNA may code for ribosomal RNAs and transfer RNAs.

replacement vectors Vectors that have a pair of insertion sites that span a DNA segment that can be exchanged with a foreign DNA fragment.

replica plating A technique to produce identical patterns of bacterial colonies on a series of petri plates. A plate containing colonies is inverted and its surface is pressed against a block covered with velveteen. The block can then be used to inoculate up to about eight additional petri plates. By marking the patterns of the colonies on the different plates with different selective properties, it is possible to identify which colonies differ in their responses to these agents.

replication fork The region of a ds DNA molecule that is unwound so that DNA replication can occur.

replication origin The site(s) on a DNA molecule where unwinding of the double helix occurs so that replication can occur. There are multiple replication origins on eukaryotic chromosomes.

reporter gene A gene used to identify or locate another gene.

repression of gene transcription The inhibition of transcription by the binding of a repressor protein to a specific site on the DNA molecule. A repressor protein is the product of a repressor gene.

response to selection (R) The difference in mean phenotypic value between the offspring of the selected parents and the mean phenotypic value of the entire parental generation before selection.

restriction endonuclease An enzyme that cuts DNA only at a limited number of specific nucleotide sequences. Also called restriction enzyme.

restriction fragment length polymorphism (RFLP) A polymorphism in an individual, population, or species defined by restriction fragments of a distinctive length. Usually caused by gain or loss of a restriction site, but could result from an insertion or deletion of DNA between two conserved restriction sites. Differences in RFLPs are visualized by gel electrophoresis.

restriction site A specific sequence of nucleotides in a piece of ds DNA which is recognized by a restriction enzyme and which signals its cleavage.

restriction site mapping DNA is digested with a series of different restriction endonucleases, the DNA fragments are electrophoresed, and the DNA fragments are ordered to produce a linear physical map of the locations of specific DNA sequences.

retroelement DNA or RNA sequences that contain a gene for reverse transcriptase. There are different classes of retroelements, including retroviruses and retrons.

retroposition The transfer of genetic information through an RNA intermediate. The genetic information carried by the DNA is transcribed into RNA, which is then reverse-transcribed into cDNA. The result is that the element is duplicated and the copy of the element is transposed.

retrosequences Retrosequences/retrotranscripts are sequences derived through the reverse transcription of RNA and subsequent integration into the genome. They lack the ability to produce reverse transcriptase.

retrotransposon A type of transposable element that transposes by means of an RNA intermediate. At least 10 families of retrotransposons are known in *Drosophila*. Often shortened to retroposon.

retrovirus RNA viruses that use reverse transcriptase during their life cycle. This enzyme allows the viral genome to be transcribed into DNA. The transcribed viral DNA is integrated into the genome of the host cell where it replicates in unison with the genes of the host. The cell suffers no damage from this relationship unless the virus carries an oncogene. If so, it could be transformed into a cancer cell. Retroviruses violate the Central Dogma during their replication. The HIV virus responsible for the AIDS epidemic is a retrovirus.

reverse genetics A particular gene is targeted for inactivation or expression in an unusual environment in order to investigate gene function. See also forward genetics.

reverse transcriptase An enzyme that synthesizes a DNA copy from an RNA template.

reverse transcription DNA synthesis from an RNA template, mediated by reverse transcriptase.

reversions Reverse mutation.

RFLP See restriction fragment length polymorphism (RFLP).

ribonuclease An enzyme that degrades RNA.

ribosomal RNA (rRNA) The RNA that acts as a structural component of ribosomes. Ribosomal RNA genes (rRNA genes) are found as tandem repeating units in the nucleolus organizer regions of eukaryotic chromosomes. Each unit is separated from the next by a nontranscribed spacer. Each unit contains three regions coding for the 28S, 18S, and 5.8S rRNAs.

ribosome A self-assembling cellular organelle made up of proteins and RNA in which translation of mRNA occurs. Ribosomes consist of two subunits, each composed of RNA and proteins. In eukaryotes, ribosome subunits sediment as 40S and 60S particles.

ribozyme An RNA molecule with catalytic activity. Ribozymes are known that self-splice

rRNA; another ribozyme is the RNA in the large subunit of the ribosome.

ring chromosome An aberrant chromosome with no ends.

RNA Ribonucleic acid, one of the two forms of nucleic acids.

RNA editing RNA editing involves altering the mRNA after transcription. This results in different proteins being produced from a single gene. The molecular mechanisms include single or multiple base insertions or deletions, as well as base substitutions. RNA editing occurs in both prokaryotes and eukaryotes.

RNA polymerase An enzyme capable of synthesizing an RNA copy of a DNA template.

RNA silencing (RNA interference) When double-stranded RNA (ds RNA) is injected into a cell, a defense response typically occurs in plants and animals in which the RNA is cut up into smaller chunks (about 22 nt long) and the fragments are then degraded. This process may be a defense against mobile DNA elements (TEs) which cause mutations when they insert themselves within or close to a gene. Experimentally, RNA interference can be used to silence cognate genes.

RNA surveillance A system in eukaryotic cells to degrade aberrant mRNAs.

RNA transcript An RNA copy of a gene.

S phase The portion of interphase in the cell cycle in which DNA replication occurs. The S phase occurs between the G_1 and G_2 phases of the interphase. Mitosis occurs after the G_2 phase.

S1 nuclease An enzyme that specifically degrades single-stranded DNAs or splits short single-stranded segments in DNA but does not attack any double-stranded molecules. Used to convert sticky ends of duplex DNA to form blunt ends or to trim off single-stranded ends after conversion of single-stranded cDNA to the double-stranded form.

satellite DNA Highly repeated DNA sequences with such a uniform nucleotide composition that, upon fractionalization of the genomic DNA and separation by density gradient centrifugation, they form one or more bands that are clearly different from the main band of DNA and from the smear created by other fragments of a more heterogeneous composition. The base composition of satellite DNA differs from that of the majority of DNA in a eukaryotic species, i.e., it is either A + T rich or G + C rich. Usually highly repetitive in sequence.

secondary transposition Movement of an element after its initial insertion into the chromosome. Secondary transposition can be induced with *P* elements in *Drosophila*.

segment polarity genes Segment polarity genes appear to determine a linear sequence of repeated positional values within each segment. Segment polarity mutants have repetitive deletions of pattern, but the deletions occur within each segment and are followed by a partial mirror-image duplication of the part that remains. Segment polarity genes (including *engrailed*$^+$, *naked*$^+$, *patched*$^+$, *wingless*$^+$, *gooseberry*$^+$, *patched hedgehog*$^+$, *porcupine*$^+$, *armadillo*$^+$, *fused*$^+$) are required either continuously or over extensive periods to maintain the segmental pattern. Most or all are required to maintain patterns in the imaginal tissues.

segmentation genes Genes, including the gap, pair-rule, and segment polarity classes of genes, that determine the number and polarity of the body segments during embryonic development in insects.

selectable marker A gene that allows identification of specific cells with a desirable new genotype. Many vectors used for genetic engineering carry antibiotic resistance genes, or other genes, that allow identification of cells containing exogenous DNA.

selection differential In artificial selection, the difference in mean phenotypic value between individuals selected as parents of the following generation and the whole population.

selfish DNA DNA that may not provide any advantage to its carrier or host but ensures its own survival. Transposable elements are considered to be selfish DNA.

semiconservative replication DNA replication in which each daughter double helix consists of one strand from the parent and one newly synthesized strand.

sensory transduction Sensory cells transform and amplify the energy provided by a stimulus into an electrical signal. Sensory transduction is probably due to a change in the ionic permeability of the sensory cell membrane, which causes a depolarization of the membrane.

sequencing The process used to obtain the sequential arrangement of nucleotides in the DNA molecule.

sericin See silk.

sex chromosome A chromosome which is involved in sex determination.

short germ band development A pattern of development found in some insects in which all or most of the metameric pattern is completed after the blastoderm stage by the sequential addition of segments during elongation of the caudal region of the embryo.

short-period interspersion pattern of genome organization This form of genome organization has single-copy DNA, 1000–2000 bp long, alternating with short (200–600 bp) and moderately long (1000–4000 bp) repetitive sequences. This pattern is found in the house fly *Musca domestica*, the Australian sheep blowfly *Lucilia cuprina*, and the wild silk moth *Antheraea pernyi*.

shotgun cloning Genomic libraries constructed from random fragments of DNA from an organism.

shotgun libraries Genomic libraries in which a random collection of a sufficiently large sample of cloned fragments of the DNA are present so that all the genes are represented.

shotgun method of transformation A method for introducing foreign DNA into cells in which tiny bullets made of tungsten or other metal are coated with DNA and shot into the cell.

silent mutation Changes in DNA that do not influence the expression or function of a gene or gene product.

silk The cocoon filament spun by the fifth-instar larva of *Bombyx mori* and other silk moths. Each cocoon filament contains two cylinders of **fibroin**, each surrounded by three layers of sericin. Fibroin is secreted by the cells of the posterior portion of the silk gland. The fibroin gene is present in only one copy per haploid genome, but these silk gland cells undergo 18 to 19 cycles of endomitotic DNA replication before they begin transcribing fibroin mRNAs. The **sericin** proteins are named because they contain abundant serines (over 30% of the total amino acids). Sericins are secreted by the cells from the middle region of the silk gland.

similarity A measure of the resemblance between two objects, usually on a scale of zero to one.

single-locus, multiple-allele model A model for sex determination in Hymenoptera.

single-strand binding proteins One of the proteins that attaches to ss DNA in the replication fork to prevent reannealing of the DNA during DNA replication.

site-directed mutagenesis Mutagenesis to produce a predetermined change at a specific site in a DNA molecule.

slot blot A hybridization technique that allows multiple samples of DNA to be applied to nitrocellulose filters in specific sites (slots) using a vacuum.

somatic cells All the eukaryotic body cells except the germ-line cells and the gametes they produce.

Southern blotting A technique developed by E. M. Southern for transferring DNA fragments isolated electrophoretically in an agarose gel to a nitrocellulose filter paper sheet by capillary action. The DNA fragment of interest is then probed with a radioactive nucleic acid probe that is complementary to the fragment of interest. The position on the filter is determined by autoradiography. The related techniques for RNA and proteins have been dubbed Northern and Western blots, respectively.

specific activity The ratio of radioactive to nonradioactive molecules of the same kind. Probes with a high specific activity can produce a more intense signal than a probe with a low specific activity.

spliceosome The RNA and protein particles in the nucleus that remove introns from premessenger RNA molecules.

ss DNA Single-stranded DNA.

stable transformation Transformation that alters the germ plasm of an organism so that the progeny transmit the trait of interest through subsequent generations.

start codon The mRNA codon, usually AUG, at which synthesis of a polypeptide begins.

stem cells Stem cells are able to self-renew and generate cell populations that differentiate to maintain adult tissues. There are about two stem cells in the ovary of *Drosophila* that maintain oocyte production.

sterile insect release method (SIRM) A genetic control technique used to control or eradicate pest insects. Large numbers of mass-produced males are given nonlethal but sterilizing doses of radiation or chemical mutagens and then released. Females in natural populations mate with the sterilized males, and produce inviable progeny. After multiple releases a new generation is not produced. Used to eradicate the screwworm from North America.

sterile male technique See sterile insect release method (SIRM).

sticky end Single-stranded ends of DNA fragments produced by restriction enzymes; sticky ends are able to reanneal.

stop codon One of the three mRNA codons (UAG, UAA, and UGA) that prevent further polypeptide synthesis. Also called termination codon.

stress proteins Also called heat shock proteins. Proteins made when the cells are stressed by environmental conditions (chemicals, pathogens, heat).

stringency Stringency, as used in hybridization reactions, refers to the conditions that can be altered to influence the ease with which a probe hybridizes to template nucleic acids.

structural gene A gene that codes for an RNA molecule or protein other than a regulatory gene.

structural genomics The study of protein structure based on DNA sequences.

subclones A DNA fragment that has been cloned into one vector may be moved, or subcloned, into a second type of vector in order to perform a different procedure.

supercoiled The coiling of a covalently closed circular duplex DNA molecule upon itself so that it crosses its own axis. A supercoil is also called a superhelix. The B form of DNA is a right-handed double helix. If the DNA duplex is wound in the same direction as that of the turns of the double helix, it is positively supercoiled. Twisting of the DNA molecule in a direction opposite to the turns of the strands in the double helix is called negative supercoiling.

symbiont An organism living with another organism of a different species.

sympatry Living in the same geographic location. Sympatric species have overlapping or coinciding distributions.

synapsis The pairing of homologous chromosomes during the zygotene stage of meiosis.

syncytium A mass of protoplasm containing many nuclei not separated by cell membranes.

synecology The study of relationships among communities of organisms and their environment.

syngamy The fusion of sperm and egg to form a zygote.

synteny Synteny refers to the fact that many genes remain grouped together in the same relative positions in the genome across taxa.

systematics The study of classification, based on evolutionary change.

tandem repeat Direct repeats in DNA codons adjacent to each other.

***Taq* DNA polymerase** A DNA polymerase that was isolated from the bacterium *Thermus aquaticus* and is tolerant of high temperatures. Used in the polymerase chain reaction.

TaqMan PCR A real-time type of PCR which uses an oligonucleotide that anneals to an internal sequence within the amplified DNA fragment. This oligo, usually 20 to 24 bases long, is labeled with a fluorescent group at its 5′ end and a quenching group at its 3′ end. When both the fluorescent and quenching groups are in close proximity on the intact probe, any emission from the reporter dye is absorbed by the quenching dye and the fluorescent emission is low. As the reaction progresses and the amount of target DNA

increases, progressively greater quantities of oligo probe hybridize to the denatured target DNA. During the extension phase of the PCR, the fluorophore is cleaved from the probe by the $5'$ to $3'$ exonuclease activity of the polymerase. Because the fluorophore now is no longer close to the quencher, it begins to fluoresce. The intensity of the fluorescence is directly proportional to the amount of target DNA synthesized and allows the researcher to quantify the reaction in "real time" without running the product on a gel.

targeted gene replacement Replacing or modifying genes in their normal chromosomal locations has not been possible with *Drosophila* until recently. The cut-and-paste model of *P*-element transposition provided a model for inserting a gene into the double-stranded gap left behind by a *P* element. The gap can be repaired, using a template provided by an extrachromosomal element that has been introduced by the investigator.

targeted gene transfer See targeted gene replacement.

targeted mutagenesis The ability to replace or modify DNA sequences in their normal chromosomal location.

TATA box See Hogness box.

taxa The general term for taxonomic groups, whatever their rank. The singular form is taxon.

taxonomy The principles and procedures according to which species are named and assigned to taxonomic groups.

tDNA-PCR Universal primers for transfer RNA can be used to generate tDNA by the PCR. The resulting fragments are visualized by gel electrophoresis and produce characteristic fingerprints for different species.

telomerase An enzyme that adds specific nucleotides to the tips of chromosomes to form telomeres.

telomere Telomeres are the physical ends of eukaryotic chromosomes. They protect the ends of chromosomes and confer stability. Telomeres consist of simple DNA repeats and the nonhistone proteins that bind specifically to those sequences.

telomere terminal transferase See telomerase.

template A macromolecular mold for synthesis of another macromolecule. Duplication of the template takes two steps; a single strand of DNA serves as the template for a complementary strand of DNA or mRNA.

termination codon One of the three codons in the standard genetic code that indicate where translation of an mRNA should stop, i.e., $5'$-UAA-$3'$, $5'$-UAG-$3'$, or $5'$-UGA-$3'$. Also called a stop codon.

thelygenic When females produce only female progeny, as in the blowfly *Chrysomya rufifacies.*

thelytoky Parthenogenesis in which no functional males are known; unmated females produce female progeny only, or rarely, a few males.

30-nm fiber Condensation of DNA in eukaryotic chromosomes involves formation of 30-nm fibers from supercoils of six nucleosomes per turn. The 30-nm fiber somehow is condensed further.

T_m The interpolated temperature along a DNA melting curve at which 50% of the duplex DNA formed in a DNA–DNA hybridization is double-stranded. The difference in T_m between homoduplex and heteroduplex curves is called ΔT_m.

tracer DNA In DNA–DNA hybridization, single-stranded single copy DNA from one species is radioactively labeled (tracer DNA) and hybridized with unlabeled DNA (driver DNA) from the same species or from different species. DNA–DNA hybridization is used to determine the degree of sequence identity between DNAs.

trailer segment A nontranslated sequence at the $3'$ end of mRNA following the termination signal, exclusive of the poly-A tail.

transcript An RNA copy of a gene.

transcription The process of producing an RNA copy of a gene.

transcriptional activator proteins Elements that stimulate transcription by binding with particular sites in the DNA.

transcriptome The transcriptome is the profile of the genes that are expressed or transcribed

from genomic DNA within a cell or tissue, with the goal of understanding cell phenotype and function. The transcriptome is dynamic and changes rapidly in response to stress or during normal cell processes such as DNA replication and cell division.

transfection Infection of bacteria with viral nucleic acid that lacks a protein coat.

transfer RNA (tRNA) A family of small RNA molecules (usually more than 50 types per cell) that serve as adapters for bringing amino acids to the site of protein synthesis on the ribosome.

transformant An individual organism produced by introducing exogenous DNA.

transformation The process of changing the genetic makeup of an organism by introducing foreign DNA. Transformation may be transient or stable (transferred to succeeding generations.)

transgene The DNA that is inserted into the genome of a cell or organism by recombinant DNA methods.

transgene suppression A variety of organisms, including insects, plants, and mammals, can inactivate multiple copies of inserted genes that overexpress proteins or are abnormally transcribed. Transgene silencing may be induced by methylation of the DNA or by posttranscriptional and transcriptional processes.

transgenic organism An organism whose genome contains genetic material originally derived from an organism (not its parents) or from a different species. The transgene(s) can be transmitted to subsequent generations (stable transformation) or can be lost subsequently (unstable transformation).

transient transformation Transient transformation involves changing the genetic makeup by introducing foreign DNA. If the genetic information is not incorporated into the germ line, the genetic changes are temporary.

transitions Transitions are point mutations that involve changes between A and G (purines) or T and C (pyrimidines).

translation The process by which the amino acid sequence in a polypeptide is determined by the nucleotide sequence of a messenger RNA molecule on the ribosome.

translational regulation Gene regulation by controlling translation. Translation of mRNA can be tied to the presence of a specific molecular signal; the longevity of a mRNA molecule can be regulated; or overall protein synthesis can be regulated.

translocation A type of mutation in which a section of a chromosome breaks off and moves to a new position in that or a different chromosome.

transovarial transmission Transmitted to the next generation through the egg.

transposable element An element that can move from one site to another in the genome. Transposable elements (TEs) have been divided into two classes, those that transpose with an RNA intermediate and those that transpose as DNA.

transposase An enzyme that catalyzes transposition of a transposable element from one site to another in a DNA molecule.

transposition The movement of genetic material from one chromosomal location to another.

transposon A transposable element carrying several genes including at least one coding for a transposase enzyme. Many elements are flanked by inverted repeats. *Drosophila melanogaster* contains multiple copies of 50–100 different kinds of transposons.

transposon tagging A method of cloning genes from *Drosophila* after they have been "tagged" by having the *P* element insert into them.

transversions Transversions are point mutations that involve changes between a purine and a pyrimidine.

triplex DNA In triplex DNA, the usual A-T and C-G base pairs of duplex DNA are present, but in addition a pyrimidine strand is bound in the major groove of the helix. DNA sequences that potentially can form triplex DNA structures appear to be common, are dispersed at multiple sites throughout the genome, and comprise up to 1% of the total genome.

ubiquitin A protein that is present in cells of both prokaryotes and eukaryotes and is highly conserved. Ubiquitin contains 76 amino acids and plays a role in proteolysis in the proteosome. Ubiquitin-conjugating enzymes add ubiquitins to proteins carrying degradation signals. The ubiquitin is recognized by proteosomes which then cut the proteins into fragments.

unique genes Genes present in only one copy per haploid genome, which includes most of the structural (protein-encoding) genes of eukaryotes.

unrooted tree A phylogenetic tree in which the location of the most recent common ancestor of the taxa is unknown.

UPGMA The use of distance measurements to group taxonomic units into phenetic clusters by the Unweighted Pair-Group Method of Analysis using an arithmetic average.

upstream Toward the 5′ end of a DNA molecule.

uracil A pyrimidine that is one of the nitrogenous bases found in RNA.

vector A DNA molecule capable of autonomous replication in a cell and which contains restriction enzyme cleavage sites for the insertion of foreign DNA.

vertical gene transfer Transfer of a gene from parents to offspring. See also horizontal gene transfer.

virus A noncellular particle that can reproduce only inside living cells; consisting only of a genetic material (either DNA or RNA) and a protein coat. Viruses are "alive" because they can reproduce, but they have no other traits of living organisms.

vitellogenin The major yolk proteins are derived from vitellogenins, which are produced by the fat body and secreted for uptake by maturing oocytes.

Western blots Proteins are separated electrophoretically, and a specific protein is identified with a radioactively labeled antibody raised against the protein in question.

wild type The normal form of an organism—in contrast to that of mutant individuals.

wobble hypothesis A hypothesis to explain how one tRNA may recognize two different codons on the mRNA. Anticodons are triplets with the first two positions pairing according to base-pairing rules. The third position "wobbles" and can recognize any of a variety of bases in different codons so that it can bind to either of two or more codons.

X chromosome A sex chromosome that is usually present in two copies in insect females (XX) and in one copy (unpaired) in males (XO or XY).

X-gal A lactose analogue (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). X-gal is cleaved by β-galactosidase into a product that is bright blue. If exogenous DNA has inserted into and disrupted the β-galactosidase gene, λ plaques will appear white or colorless. Plaques without recombinant vectors will be blue.

Y chromosome A sex chromosome that is characteristic of males in species in which the male typically has two dissimilar sex chromosomes (XY).

Z chromosome One of the sex chromosomes found in heterogametic ZW female insects.

Z-DNA A structural form of DNA in which the two strands are wound into a left-handed helix rather than a right-handed form.

zinc finger protein Proteins with tandemly repeating segments that bind zinc atoms. Each segment contains two closely spaced cysteine molecules followed by two histidines. Each segment folds upon itself to form a fingerlike projection. The zinc atom is linked to the cysteines and histidines at the base of each loop. The zinc fingers serve in some way to enable the proteins to bind to DNA molecules, where they regulate transcription.

zygote A fertilized egg formed as the result of the union of the male and female gametes.

INDEX

B